McKinsey Global Institute
McKinsey Center for Government
McKinsey Business Technology Office

October 2013

Open data: Unlocking innovation and performance with liquid information

The McKinsey Global Institute

The McKinsey Global Institute (MGI), the business and economics research arm of McKinsey & Company, was established in 1990 to develop a deeper understanding of the evolving global economy. Our goal is to provide leaders in the commercial, public, and social sectors with facts and insights on which to base management and policy decisions.

MGI research combines the disciplines of economics and management, employing the analytical tools of economics with the insights of business leaders. Our "micro-to-macro" methodology examines microeconomic industry trends to better understand the broad macroeconomic forces affecting business strategy and public policy. Current research focuses on four themes: productivity and growth; the evolution of global financial markets; the economic impact of technology and innovation; and urbanization.

For further information about MGI and to download reports, please visit www.mckinsey.com/mgi.

The McKinsey Center for Government

The McKinsey Center for Government (MCG) is a dedicated center of expertise and innovation within McKinsey's Public Sector Practice to support government leaders around the world in their goal to achieve greater and lasting performance for the citizens they serve. MCG's mission is to help public-sector leaders understand the forces that are driving change, design the best strategies for success, develop the tools and capabilities needed for effective execution, and deploy these tools for impact.

The center focuses on the critical and common challenges that governments face and creates opportunities for government leaders to learn from successful experiences, innovations and different contexts. Drawing on a network of external experts and McKinsey practitioners from across the public, private, and non-profit sectors, MCG provides governments with both new and proven knowledge and tools to confront critical challenges in a context of limited resources and creates opportunities for governments to learn from one other. For more information, please visit www.mckinsey.com/client_service/public_sector/mckinsey_center_for_government.

The McKinsey Business Technology Office

The McKinsey Business Technology Office (BTO) was created in 1997 to serve executives on their information technology needs in multiple sectors including automotive, energy, finance, health care, high tech, media, mining, pharmaceuticals, public sector, retail, telecommunications and transportation.

Operating in 59 cities in 27 countries, BTO's experts help clients in the following areas: optimizing enterprise architecture, infrastructure, and application management; delivering critical IT projects on time and within budget while capturing maximum business benefit and value; leveraging technology to digitize business processes; developing IT strategies that are tightly aligned to business goals, and designing corresponding organization and governance approaches that deliver the highest value; and understand the long-term technological forces at work that have an impact on their business. For more information, please visit www.mckinsey.com/client_service/business_technology.

Copyright © McKinsey & Company 2013

McKinsey Global Institute
McKinsey Center for Government
McKinsey Business Technology Office

October 2013

Open data: Unlocking innovation and performance with liquid information

James Manyika
Michael Chui
Peter Groves
Diana Farrell
Steve Van Kuiken
Elizabeth Almasi Doshi

Preface

It has long been recognized that free markets and free societies thrive on the free exchange of information. Transparency of market prices and of government operations are the foundations of efficiency and integrity, encouraging participation by market players and citizens, and spurring new ideas and innovations. Today, with massive computing power and data analytics, it is possible to make open digital information "liquid" and shareable to an unprecedented degree.

Open data—from both public and private sources—are adding a new dimension to big data analytics and giving rise to novel, data-driven innovations. Businesses are finding new ways of segmenting markets by blending open data with proprietary data and discovering new ways to raise productivity by using open and proprietary data to benchmark operations. Consumers are benefiting from open data by gaining more insights into what they buy, where they go to school, and how they get around (for example, with mobile apps that use open data to show the flow of traffic and public transit).

In this paper we identify ways in which open data can create economic value, both in terms of revenue and savings and in economic surplus, such as the time savings that commuters gain when they avoid congestion. We estimate potential annual value that use of open data could bring in seven domains: education, transportation, consumer products, electric power, oil and gas, health care, and consumer finance. The estimates are intended to be indicative and not exhaustive. Nor do we attempt to estimate the value of all of the considerable societal benefits that can be derived from use of open data. Our hope is that this work will inform the agenda for adopting and managing open data in both the public and private sectors and provide a lens for examining the critical issues about privacy and protection of proprietary information that need to be resolved before the full value of open data can be realized.

This report is the result of collaboration between the McKinsey Global Institute, McKinsey's Global Public Sector Practice, and McKinsey's Business Technology Office (BTO). The effort was led by James Manyika, an MGI director, and Michael Chui, an MGI principal; Diana Farrell, a director of the McKinsey Center for Government; and Steve Van Kuiken, a BTO director, and Peter Groves, a BTO principal.

The project team was led by Elizabeth Doshi and Dieter Kerkfeld. The research team consisted of Aditya Dada, Mike Hodel, Ryan Ko, Lea Thiel, Maria Fernandez Vidal, Adam Wiman, and Katie Wynbrandt. Geoffrey Lewis provided editorial support, and we thank Julie Philpot, Marisa Carder, Rebeca Robboy, and Gabriela Ramirez from MGI for their help in designing and producing this report.

McKinsey experts who guided our research in five of the seven domains are Jimmy Sarakatsannis in education, Simon Kennedy in transportation, Julien Boudet in consumer products, Mike Greene in electricity, and Jim Nolten in oil and gas. Our research in the health-care applications of open data is based on the work of Peter Groves, Basel Kayyali, David Knott, and Steve Van Kuiken. In consumer finance our work was guided by Murdoch Martin in real estate, Jonathan Steitz in retail banking, and Lori Sherer in insurance. We thank Joi Danielson and Shalabh Gupta for their initial research, which shaped the study.

We would also like to thank our McKinsey colleagues who contributed their knowledge in of industries and sectors that we examined. Education: Shaistah Bahrainwala, Susan Colby, Adam Eichner, Tom Isherwood, Ali Jaffer, Sarah Kramer, Andy Moffit, and Doug Scott; transportation: Tyler Duvall, Josh Kowitt, John Means, Jürgen Müller, Jules Seeley, Yakov Sergienko, and Alice Woodwark; consumer products: Jim Brennan, Pat Callinan, Javier Cazanga, Brian Henstorf,

Shruti Lal, Ross McMillan, Lou Rassey, Roger Roberts, Shawn Stewart, Rob Turtle; electricity: Priya Atur, Shannon Bouton, Katrin Heisterberg de Gimenez, Mike Linders, Jessica O'Connor-Petts, Manya Ranjan, Frank von Willert, and Alex Xiong; oil and gas: Jennifer Brady, Tim Fitzgibbon, Pat Graham, Stuart Morstead, Kunal Patel, Joe Quoeyeser, Tor Jakob Ramsoy, and Asli Ucyigit; consumer finance: Marcy Ellis-Williams, Antonio Grimaldi, Rob Hayden, Ayan Sarkar, Robert Schiff, Andrew Sellgren, Johnson Sikes, Hiek van der Sheer, and David Yates.

This work has benefited greatly from the generous advice of our external advisers, Martin Baily, senior fellow in the economic studies program and the Bernard L. Schwartz chair in economic policy development at the Brookings Institution, Karen Kornbluh, senior fellow for digital policy at the Council on Foreign Relations, and Hal Varian, professor emeritus, information sciences, business, and economics at the University of California at Berkeley. We also thank Todd Park, the chief technology officer of the United States; Beth Simone Noveck, founder and director at the Governance Lab at New York University; and Nick Sinai, US deputy chief technology officer, for their insights.

In addition, we thank Lauren Dyson and the Code for America fellows for their guidance, as well as Stacy Donohue, director of investments, Omidyar Network; Derek Eder, open data Web developer; Sara Ittelson, director of business development, Knewton; Ian Kalin, director of open data for Socrata; Pallavi Kaushik, technology, innovation, and early-stage development, Motorola Solutions; Ted Mitchell, president and chief executive officer of NewSchools Venture Fund; Daniel X. O'Neil, executive director, Smart Chicago; Tom Schenk, director of analytics and performance management, City of Chicago; and Trond Unneland, vice president and managing executive, venture capital, Chevron Technology Ventures.

This report is part of our ongoing work on the impact of technology on the economy. Our goal is to provide a fact base and insights about important technological developments that will help business leaders and policy makers develop appropriate strategies and responses. As with all of MGI's work, this report has not been sponsored in any way by any business, government, or other institution.

Richard Dobbs
Director, McKinsey Global Institute
London

James Manyika
Director, McKinsey Global Institute
San Francisco

Jonathan Woetzel
Director, McKinsey Global Institute
Shanghai

Diana Farrell
Director, McKinsey Center for Government
Washington, DC

Steve Van Kuiken
Director, McKinsey Business Technology Office
New Jersey

October 2013

More open data for more users...

40+

Number of countries with government open data platforms*

90,000+

Data sets on data.gov (US site)*

1.4 million

Page views for the UK open data site in the summer of 2013

102

Cities that participated in 2013 International Open Data Hackathon Day

1 million+

Data sets made open by governments worldwide

* As of 2013

... can lead to more value

$3 trillion

Approximate potential annual value enabled by open data in seven "domains"

3 billion

Metric tons of carbon dioxide equivalent emission reductions from buildings that could be identified through the use of open data

35

Hours per year could be saved by commuters from schedule changes based on open data

100,000+

Medical, health, and fitness apps for smartphones

50%+

Consumer share of potential value of open data

Contents

Executive summary 1

1. Open data in education 17

2. Open data in transportation 31

3. Open data in consumer products 43

4. Open data in electricity 57

5. Open data in oil and gas 69

6. Open data in health care 79

7. Open data in consumer finance 91

Bibliography 101

Executive summary

The global economy is beginning to operate truly in real time, with constant streams of data showing where consumers are shopping, ships are traveling, energy and money are flowing. Big data analytics make it possible to work through massive amounts of real-time and historical information to find unseen patterns and discover anomalies that can indicate opportunities for new products and services and new ways of operating more efficiently. Now, a complementary trend is under way. Open data—the release of information by governments and private institutions and the sharing of private data to enable insights across industries—provides additional depth to big data applications and makes possible entirely new ones, such as the smartphone app that tells commuters when the next bus will arrive. This trend has profound implications for companies, governments, and individuals.

Many open data initiatives, particularly in the public sector, have been motivated by societal goals such as improving the transparency and accountability of institutions, and much has been written about the importance of these efforts. Our research focuses on economic value that can be created by open data. Making data more "liquid" (open, widely available, and in shareable formats) has the potential to unlock large amounts of economic value, by improving the efficiency and effectiveness of existing processes; making possible new products, services, and markets; and creating value for individual consumers and citizens.[1] Realizing this potential will involve creating safeguards for personal privacy and business confidentiality, investments in technology, and changes in mindsets and work processes.

The value of making information open and available has long been recognized. Thomas Jefferson, the third president of the United States, understood that the development of the economy and the health of the democracy depended on the free flow of information. Jefferson funded the expedition of Lewis and Clark in 1803 to gather all the information they could on the vast wilderness that had been acquired from France in the Louisiana Purchase—providing data on topology, human settlements, rivers, minerals, soil, flora, fauna, and weather that enabled the settlement of the frontier and the rapid growth of the economy. Governments since have gathered and shared (to varying degrees) vital information: GPS data, weather data, and census information are examples of information sets that are collected by public agencies in the course of their work and then made freely available for use by citizens, businesses, and academics. Open data brings this principle into the 21st century and vastly extends the power of shared information by making available massive, fine-grained, and timely data, which through advanced computing and analytics yield novel insights. Entrepreneurial companies are seizing the opportunity: Climate Corporation, a startup that was recently acquired for about $1 billion, combines 30 years of weather data,

1 We use the terms "open data" and "liquid data" interchangeably.

60 years of crop yield data, and 14 terabytes of soil data—all from government agencies—for such uses as research and pricing crop insurance.[2]

We are studying the economic impact of open data now because this is a critical moment, when demand for data-driven insights intersects with more data being made open, and there are rapid advances in analytic capabilities, which accompany the adoption of big data. Although the open data phenomenon is in its early days, we see clear potential to unlock significant value by applying advanced analytics to a combination of open and proprietary data. As data are made more liquid, individuals and organizations can take advantage of the data to create value. The breadth and diversity of information in open sources could make open data a highly cost-effective source of critical insights in many markets.

Put to best use, liquid data can also become a critical element for breaking down information gaps across industries, sharing insights that can raise productivity, enable innovation, and replace traditional and intuitive approaches with data-driven processes. Analytics powered by open data can also help uncover consumer preferences, anomalies in costs, and variations in performance—all of which can inform new products and processes. However, these benefits must be weighed against privacy concerns, as well as loss of control over information about oneself or one's company. And, even if the data are free, there are costs associated with the effort to measure, analyze, and incorporate insights from the data into daily decisions for both consumers and businesses.

This research focuses on quantifying the potential value of using open data in seven "domains" of the global economy: education, transportation, consumer products, electricity, oil and gas, health care, and consumer finance. We identify the "levers" through which open data can create economic value and also explore the barriers to adoption and "enablers" for capturing value by making data more open.

Based on this analysis, we estimate that open data have the potential to enable more than $3 trillion in additional value annually across these domains.[3] This represents our estimate of true economic growth, but it does not account for share shifts that are likely to occur within industries and between consumers and business. Value can arise in a number of ways, including equipping workers with the skills to raise productivity, allowing marketers to micro-segment populations more successfully, and boosting performance across segments by sharing benchmarks, market data, and best practice information. Consumers stand to gain by saving money through greater price transparency and using more information to make decisions.

While use of open data varies by domain, there are common themes in how value is created. Techniques such as segmenting populations and exposing variability, which are commonly used in big data analytics, apply to open data analytics as well. In many instances, consumers will reap the greatest benefits from open

2 David Kesmodal, "Monsanto to buy Climate Corp. for $930 million," *Wall Street Journal*, October 2, 2013.

3 Throughout this report we express value in terms of annual economic surplus in 2013 US dollars, not the discounted value of future cash flows; this valuation represents estimates based on initiatives where open data are necessary but not sufficient for realizing value. Often, value is achieved by combining analysis of open and proprietary information to identify ways to improve business or government practices. Given the interdependence of these factors, we did not attempt to estimate open data's relative contribution; rather, our estimates represent the total value created.

data applications, usually by gaining access to information that allows them to make better decisions. Another common theme across the domains is risk. Open data creates new risks for companies and place new demands on government to protect privacy and prevent misuse of information.

DEFINITIONS AND EXAMPLES OF OPEN DATA SETS

While it has always been possible to share information openly, until information was in a digital format that could be easily shared and analyzed, the potential for open data was limited. Since the 1970s, however, governments have been able to share more data and more kinds of data electronically. More recently, private players—established companies and emerging data-sharing specialists—have also begun to make their data available to others. These open data sets share the following characteristics[4]:

- **Accessibility:** A wide range of users is permitted to access the data.
- **Machine readability:** The data can be processed automatically.
- **Cost:** Data can be accessed free or at negligible cost.
- **Rights:** Limitations on the use, transformation, and distribution of data are minimal.

Data sets range from completely open to completely closed across these four dimensions. In Exhibit E1 we see how data are open or closed based on the four characteristics that define open data.

Exhibit E1
How data are open or closed, based on four characteristics

Degree of access	Everyone has access	Access to data is to a subset of individuals or organizations
Machine readability	Available in formats that can be easily retrieved and processed by computers	Data in formats not easily retrieved and processed by computers
Cost	No cost to obtain	Offered only at a significant fee
Rights	Unlimited rights to reuse and redistribute data	Re-use, republishing, or distribution of data is forbidden

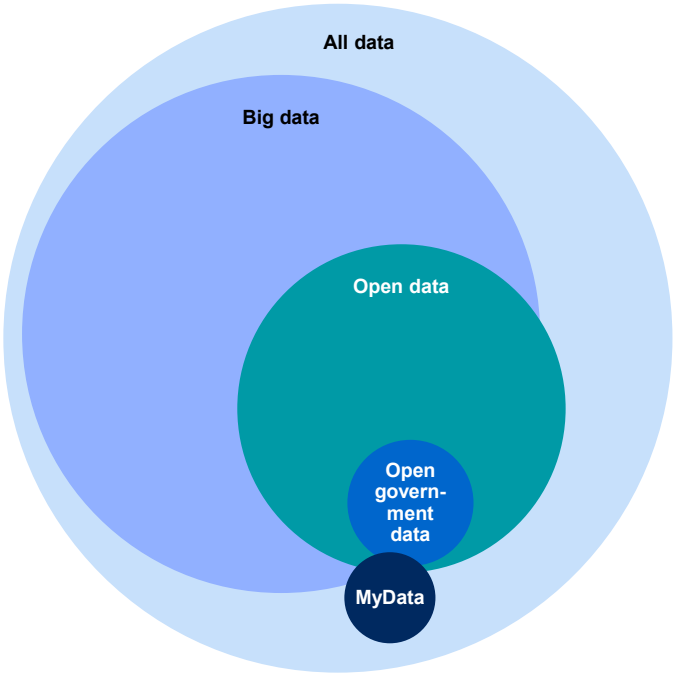
SOURCE: McKinsey Global Institute analysis

4 This is similar to the definition put forth in "8 principles of Open Government," which emerged at an early meeting on open data in Sebastopol, California, in December 2007. Four of those principles (data must be complete, data must be primary, data must be timely, and data must be non-proprietary) describe properties of open data that make it more valuable. However, we do not include them in our definition because we believe data can still be classified as open or liquid, even if they do not meet these criteria.

Open data sets also are defined in relation to other types of data, especially big data (Exhibit E2). “Big data” refers to data sets that are voluminous, diverse, and timely. Open data is often big data, but “small” data sets can also be open. We view open and big data as distinct concepts. “Open” describes how liquid and transferable data are, and “big” describes size and complexity of data sets. The degree to which big data is liquid indicates whether or not the data are open.

Exhibit E2
How open data relates to other types of data

ILLUSTRATIVE



SOURCE: McKinsey Global Institute analysis

Open data sets—whether big or small—can come from the government or other institutions and enterprises, and from individuals. Open data initiatives in the public sector, in which governments release data, are some of the most prominent examples of this trend, but we do not consider open data to be synonymous only with data released by governments.

Finally, the open data concept is associated with “MyData,” which involves sharing information collected about an individual (or organization) with that individual. For example, some hospitals now provide individual patients with access to their own medical records data. Providing aggregate statistics (a form of open data) alongside MyData allows useful comparisons; some utilities show consumers how their energy use compares with that of neighbors to encourage conservation. In some cases, individuals are allowed to modify or correct the data provided to them, which improves the quality of the data.

Open data sets vary in scope and source. They can be local, national, or global and can be obtained from both government and commercial sources (see Box 1, “An open data sampler”).

Box 1. An open data sampler

- **New York City building utilities benchmarking.** Detailed information on energy and water consumption for each non-residential building in New York City was released in 2011 and is used by building operators to benchmark the energy efficiency of their buildings and identify opportunities for improvements.
- **Real-time train movements.** Trafikverket, the transportation agency in Sweden, publishes real-time data on train departure and expected arrival times and track numbers for all trains traveling through the country. Third parties have used these data to create applications that allow travelers and shippers to make better-informed decisions on travel modes and routes.
- **Available parking spots.** Real-time open data about available parking locations has been made available in cities such as Singapore, Chicago, and San Francisco. Applications that use this open data help drivers locate parking spaces, reducing parking search time. These data can also be used in infrastructure planning.
- **Census.** Census data are a classic example of open data. In the United States, the federal government must compile and publish census data periodically, as stipulated in the Constitution. The US census provides detailed information on demographic and socioeconomic trends, down to the zip code level, helping government guide delivery of services (for example, locating schools) and enabling stores to customize formats and merchandise. Other countries release similar information.
- **Social media entries.** Social media are a growing source of wide-ranging information on customer preferences and experiences. Access to the full stream of social media content from a particular platform often requires some kind of commercial arrangement, so it is not fully open along the cost dimension, but it is relatively liquid.

OPEN DATA INITIATIVES

More than 40 countries—from every region of the world and at every stage of development—have established open data initiatives. These nations are opening up all kinds of data sets to promote economic development, spark innovation, and find ways to make government work better. India has released 3,500 data sets, mostly of agricultural information. Singapore has shared 8,600 data sets from 60 public agencies. The World Wide Web Foundation's Open Data Index rates governments on 14 data-sharing metrics and in 2012 ranked the United States, Mexico, Singapore, the United Kingdom, and New Zealand as the top five most open governments.

In the United States the opening of government data accelerated after President Obama, on his first day in office in 2009, signed an executive order stating that all government information that did not have to be kept secret for security or privacy reasons should be made public. The administration also launched the Open Data Initiative to publish government data and the data.gov website to distribute the data, which has grown from 47 "open" data sets in March 2009 to more than 90,000 sets covering everything from energy use to consumer complaints.[5] The United Kingdom began its open data effort in September 2009 with 2,500 data sets, which grew to more than 10,000 data sets in 2013.[6] Kenya, which in 2011 became the first sub-Saharan African nation to launch an open data initiative, claims that opening up government procurement data and exposing price

5 Vivek Kundra, *Data.gov: Pretty advanced for a one-year-old*, Data.gov.us, October 2013.

6 "Tim Berners-Lee unveils government data project," BBC World News, January 21, 2010.

differences can save the government $1 billion annually.[7] The commitment to making data more open was re-affirmed at the June 2013 G8 summit through the Open Data Charter, which establishes "an expectation" that the default policy should be that all government data be published openly. The charter also notes that there are legitimate reasons why some data cannot be released.

State and local governments are also creating platforms to release their data. In the United States, California and Texas have identified millions of dollars a year in savings by releasing budgetary information and enabling citizens to spot potential opportunities to cut costs. To improve transparency and citizen engagement, cities such as Boston, Buenos Aires, Dublin, Helsinki, and New York have made significant amounts of information public, from restaurant health inspection scores to school performance ratings.

While governments are natural sponsors of open data initiatives—both as sources of data and as regulators of open data use—other institutions and enterprises are also beginning to release their data and invest in programs that rely on the use of open data. In Japan, for example, citizens mobilized to create an open database of radiation readings in different parts of the country following the Fukushima earthquake. Data generated by the automatic identification system (AIS) used by ships, which shows the location of ships at sea for collision avoidance and emergency rescue, have been opened, allowing shippers to see the locations and destinations of all larger commercial ships so they can judge which ports are likely to be congested. A growing industry of third-party brokers is scouring the world's data assets, making data more liquid by aggregating, integrating, and selling access to data (some proprietary, some open), sometimes along with analytics services.

KEY FINDINGS

We focused our analysis of how use of open data can create economic value globally in seven domains.[8] The domains were selected to give a broad range of examples of how open data can create value in the public and private sectors. These are not necessarily the domains in which open data could have the most impact, but were selected to show a wide range of examples. The following are key findings from our research.

- **An estimated $3 trillion in annual economic potential could be unlocked across seven domains.** These benefits include increased efficiency, development of new products and services, and consumer surplus (cost savings, convenience, better-quality products). We consider societal benefits, but these are not quantified. For example, we estimate the economic impact of improved education (higher wages), but not the benefits that society derives from having well-educated citizens. We estimate that the potential value would be divided roughly between the United States ($1.1 trillion), Europe ($900 billion) and the rest of the world ($1.7 trillion).

7 Elana Berkowitz and Renée Paradise, "Innovation in government: Kenya and Georgia," *McKinsey Quarterly*, September 2011.

8 The analysis of health care is based on previous McKinsey research published in *The "big data" revolution in health care: Accelerating value and innovation*, McKinsey Center for US Health System Reform and McKinsey Business Technology Office, January 2013. Our work in the consumer finance domain highlights some key uses of open data but does not size the potential value comprehensively.

- **Open data enhances the value potential of big data analytics and provides additional opportunities.** Often the value from open data is realized by combining open and proprietary data. The additional open information enhances the opportunities from the five big data levers identified by MGI.[9] The levers are: creating transparency to unearth information to make better personal, business, and governing decisions; exposing variability and enabling experimentation to identify areas for improvement; segmenting populations to tailor actions such as creating custom marketing offers; augmenting or automating human decision making; and defining new products, services, and business models. About one-third of the estimated potential value from open data comes from benchmarking, an exercise that exposes variability and also promotes transparency within organizations. In addition to creating value with the big data levers, open data creates value in its own ways. When entrepreneurial companies and established businesses gain access to existing data (often government data) that were previously unavailable, they can significantly refine the process of defining and creating products and services. Open data's transparency also affects markets, fostering competitiveness by making more information available and creating opportunities to better match supply and demand. Liquid data also enables new opportunities for large-scale collaboration among individuals, companies, governments, and other organizations. Finally, open data can enhance the accountability of institutions such as governments and businesses and can raise the quality of decision by giving citizens and consumers more tools to scrutinize business and government.

- **Consumers stand to gain the most.** Consumers are already beginning to benefit from open data through price transparency (for example, by using online shopping sites that offer price comparisons). Other information about products and services could be made available through open data (e.g., whether trains are running on time or the labor and environmental practices of manufacturers) and could be used by consumers to select the products and services that best match their preferences. Opening MyData gives consumers better visibility into their own consumption, often revealing information that can lead to changes in behavior. Open data also gives individuals (as consumers and citizens) new channels to provide input to improve the quality of goods and services (including public services) and the quality of data. Together, more than 50 percent of the value potential we estimated is in consumer and customer surplus.

- **Open data helps businesses raise productivity, and create new products and services.** Using open data can help companies improve the productivity of current business processes and can lead to new products, services, and entire lines of business for both established companies and entrepreneurs. Open data can also help create more tailored products and services. By sharing data openly (yet anonymously), companies can benchmark performance and share best practices. There is also an open data industry, composed of companies that aggregate and sell data and advise companies on open data use.

9 See *Big data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute, May 2011.

- **Open data creates new risks, including threats to reputation and loss of control over confidential information.** Open data can expose individuals and businesses to several risks, especially reputational ones. Opening information such as electricity use or school performance to create aggregated views of population behavior raises serious concerns among consumers who fear that their data will be tied to them and could harm their economic or social standing. Conceivably, credit card companies could raise interest rates on households that waste electricity or the inadvertent release of information about a particular student could lead to bullying. For businesses, open data released by third parties could expose poor environmental or labor practices or show that their products or services compare poorly for price and quality. Companies can also put consumers off by using open data to create online advertisements or marketing offers that show that the company knows too much about the consumer. Another risk arises in sharing benchmarking data among businesses, if the pooled data inadvertently reveal confidential information.

- **Governments have a central role to play as a source of open data and as a regulator.** Government could use policy and other mechanisms to determine the nature of open data within society. To promote open data approaches, government could assuage concerns of both consumers and businesses about the safety of open data and help educate the public about the potential benefits to the economy and society. Governments can use their influence to make data more open, through dialogue as well as regulation, while creating policies that thoughtfully address issues such as privacy, confidentiality, intellectual property protection, and liability. Government agencies also hold vast stores of data, which, if made open, could enable the creation of large amounts of economic value. Making government data available not only enables value creation, but also sets the tone for openness among other institutions.

- **Making data more liquid is necessary but not sufficient for capturing value of open data.** Once data are open, other developments and actions are needed to realize value potential. A vibrant ecosystem of developers will be necessary to transform open data into valuable tools. Thoughtful policies that protect intellectual property and ensure privacy and confidentiality will be needed to give consumers and institutions confidence to move forward with open data. Policies will need to be refined continually to balance the value of anonymous, aggregated information about an individual with the increased risk of identification. Investment in technology is needed to collect data, create suitable platforms for sharing data, and perform analyses to uncover valuable insights. Standards can be developed that will make data from multiple sources comparable. Releasing metadata (data about data) can make open data more usable. An open data marketplace may be needed to provide clear channels for sharing liquid data and to build a community with group norms and rules, which would discourage malicious manipulation of data through economic and social pressure. Third-party services will likely play an essential role in cataloging, cleaning, and parsing information that is not released in machine-readable format. Organizations will also need to acquire and develop the talent, processes, and cultures to complement their technological investments. Key skills include the ability to perform analyses, create useful reports and tools based on open data, and incorporate data into managerial decision-making processes.

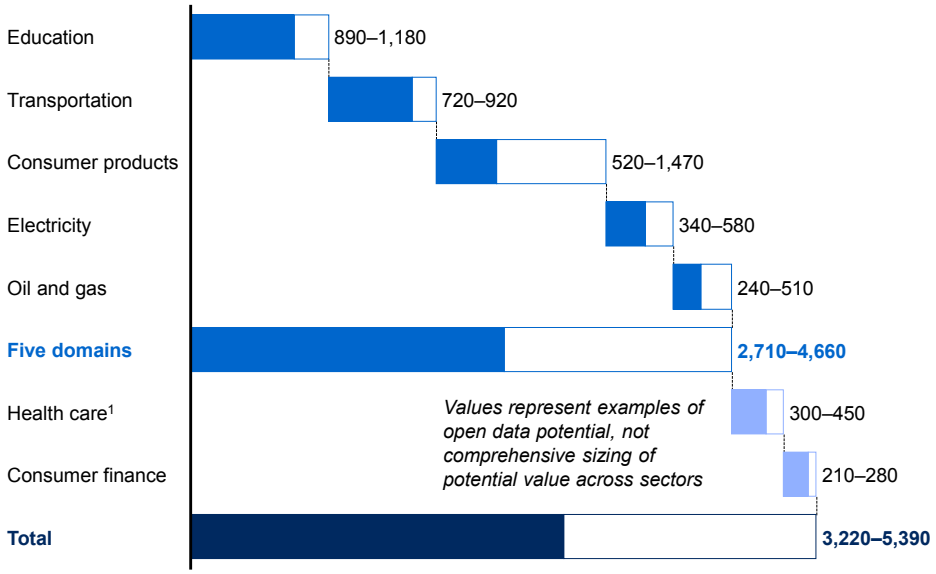
OPEN DATA IN SEVEN DOMAINS

To understand the potential economic value that might be expected from widespread use of open data, we looked at seven areas of the global economy. For each, we identified the most important levers for unlocking value with open data and we estimated how much annual value to the economy each lever might help enable (Exhibit E3). These levers work by enabling better decision making, such as spurring energy efficiency by revealing relative consumption; by providing the insights for customized products and services, such as personalized learning programs; or by exposing anomalies in performance data that lead to better processes, such as using public budget data to find opportunities to save on procurement. In addition to the potential economic benefits to both the public and private sectors that we describe here and in the following chapters, we recognize that there are separate and potentially powerful societal benefits from making data about government more open (see Box 2, "The societal benefits of open data").

Exhibit E3

Open data can help unlock $3.2 trillion to $5.4 trillion in economic value per year across seven "domains"

$ billion



1 Includes US values only.
NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

Box 2. The societal benefits of open data

The open data phenomenon began with the release of government data and it is still often associated with its potential impact on government. Many of the early successes of open data resulted from government release of information that led to greater accountability within government, such as exposing tax fraud in Canada and monitoring of corruption in Brazil after public records were made broadly accessible.[1]

Opening up public data and arming the public with the information to make concrete policy recommendations has also improved citizen engagement and participation. Peng Shi, a graduate student, created the new methodology for public school assignments in Boston, using information released by the city on the quality and location of schools.[2] His algorithm, which shows parents choices based on school quality and distance, was chosen over five other plans that had been developed through more traditional channels. Open data can also be used to monitor the use of public goods and encourage optimized investments of public goods.

Open data also enables collaborations across sectors in both public and private settings. Following the earthquake that struck Haiti in 2010, volunteers across the world combined data from sources such as satellite maps, World Health Organization maps of health facilities, and locations of police facilities from the Pacific Disaster Center, and data from mobile phone companies. The OpenStreetMap project became a critical source of reliable information to guide both governments and private aid workers and supplies to hospitals, triage centers, and refugee camps. This map helped responders better match the supply and demand for various resources in the aftermath of the disaster, significantly improving the services delivered.

1 Marcos Cruz and Alexandre Lazarow, "Innovation in government: Brazil," *McKinsey Quarterly*, September 2012.

2 James Vaznis, "MIT has plan for Boston school assignments," *Boston Globe*, October 28, 2012.

Education

Using open data in education could enable approximately $890 billion to $1.2 trillion in value annually. The largest potential benefit comes from using open data to improve instruction by identifying the most effective strategies and tools for teaching specific skills and knowledge; students who acquire higher skills can expect higher lifetime earnings. Open performance data can also be used by students and parents to make more informed decisions about choices of schools and academic or vocational concentrations. Barriers include privacy concerns, especially for student and teacher performance data, and concern around the loss of autonomy of teachers in the classroom.

Transportation

We estimate that $720 billion to $920 billion in value per year could be enabled by the use of open data in transportation. The greatest potential source of value is increased productivity and time saving for individuals from using open data to reduce travel times. Additional value can be gained by using open data to improve the efficiency of public transportation and freight operations, through adjusting train and bus schedules to better match demand and optimizing operations based on industry-wide benchmarks. Open data can also inform infrastructure investments. To capture much of the consumer value of open data in transportation, consumers will need access to easy-to-use applications that provide real-time location and estimated travel times of different transportation options.

Consumer products

We estimate that $520 billion to nearly $1.5 trillion in value per year could be enabled through the use of open data in consumer products. Consumers could benefit from price transparency, as well as access to additional open data about products (the provenance of packaged food, for example) and suppliers (such as environmental and labor practices). Open data can help manufacturers and retailers generate sales and increase loyalty, by more finely segmenting consumer types and customizing store layouts and assortments for specific neighborhoods. Manufacturers can harness open data in the form of social media chatter or "likes" to inform product design decisions and improve after-sale interactions. While consumers stand to gain the most from open data, privacy concerns represent a very significant potential barrier.

Electricity

Use of open data in the generation, distribution and consumption of electricity could help bring about value of as much as $340 billion to $580 billion annually. Significant investments in technology and operations will be required to capture the value from conservation and improved operations. By providing consumers with detailed data about their energy consumption and showing how other similarly situated consumers (or businesses) use electricity, customers can discover energy-efficiency opportunities. Utilities can also benefit from sharing benchmarking data to improve project management (e.g., streamlining permitting processes) and operations. Utilities and regulators will need to develop effective programs to use open data, build trust among consumers by establishing clear privacy protections, and encourage organizational cooperation through data-sharing schemes.

Oil and gas

Across the oil and gas value chain, open data could help enable $240 billion to $510 billion a year in value, by improving investment decisions about where to explore for new reserves and build downstream facilities. Openly sharing benchmarks can improve investing processes and operations. Sharing consumption data can help consumers make better-informed decisions about energy use (reducing natural gas consumption, for instance). Many benefits depend on companies sharing data that has been regarded as proprietary. Therefore, the business cases and requirements for sharing information will need to be made clear.

Health care

Previous McKinsey research examined the potential impact of more liquid data in health care and identified $300 billion to $450 billion in annual value that could be unlocked in the United States alone.[10] Potential sources of value include enabling people to take an active role in disease prevention and treatment; helping providers determine what is the most timely, appropriate treatment for each patient; matching patients with the most appropriate providers; ensuring the cost-effectiveness of care; and identifying new therapies and approaches to delivering care. Capturing this value will require changes in how care is delivered and paid for (in the United States), standards for data governance and usability, persuading

10 *The "big data" revolution in health care: Accelerating value and innovation*, McKinsey Center for US Health System Reform and McKinsey Business Technology Office, January 2013.

providers to share data, and investing in the capabilities of all players to make the most of open data.

Consumer finance

We examined the potential impact of open data in three areas of consumer finance: banking, insurance, and real estate. Though the potential impact in these three areas is not sized comprehensively, we estimate that $210 billion to $280 billion can be captured annually from a few of the largest examples we consider. In banking and insurance, there are significant opportunities to increase value through use of open data to improve product design and underwriting. In particular, open data can be used to assess risks for consumers who do not have a credit history, opening up a large potential source of new business (worldwide, half of adults have no banking relationships). Open data can also be used in fraud prevention and detection: affiliations revealed in social media connections have been used to uncover insurance fraud rings. Consumers can also use open data to choose among a confusing array of complex financial products. In real estate, open data can help match buyers and renters with properties, and inform development and infrastructure investments. Barriers to realizing the value of open data in consumer finance include substantial concerns about privacy by consumers and a lack of legal frameworks regarding the use of new forms of data in banking and insurance.

IMPLICATIONS FOR STAKEHOLDERS

Governments, companies, and individuals will need to understand how to take advantage of open data. All stakeholders—governments, non-profits, businesses, individuals (as consumers and citizens)—have roles to play in maximizing the benefits of open data. Deriving valuable insights from open data will require new rules and procedures and new attitudes as well as investments in technology and capabilities.

- **Governments.** Governments can set the tone for open data within a society, both by releasing data and shaping the policy environment. Public-sector agencies can be a key source of open data. An important first step is to set priorities for data release that are based on potential value, rather than ease of "opening" the data for sharing. Agencies can establish clear rules to govern the type of data that should—or should not—be released, with particular focus on safety, national security, privacy, liability, intellectual property rights, and confidentiality. Government leaders can champion the focus on open data across agencies and help make sure that the default decision is to release data whenever possible. As primary sources of open data, governments at all levels can be leaders in developing easy-to-use platforms for accessing open data. Leaders can also seek public-private partnerships or collaborations to support open data activities. For instance, the United States Patent and Trademark Office has worked with Google to post information such as awards and published applications in an easily searchable format.

  As the maker of laws and enforcer of regulations, government can shape the legal and economic environment that maximizes the potential societal value from the use of open data, while addressing the legitimate privacy and intellectual property concerns of individuals and organizations. This may include policies on who can access information, the type of information that can be collected or used in certain functions, and protocols for notifying

individuals when information about them is made available. Elected officials and policy makers can convene stakeholders to discuss innovative ideas, set standards, and make data accessible and usable. Regulators can also create policies to encourage companies to collect and release data (with appropriate protections). Additionally, public-sector agencies can help address the need for workers with strong data and analytical skills through education and immigration policies.

- **Entrepreneurs.** Open data creates opportunities for entrepreneurs (or in-house "intrapreneurs" in corporations) to create new revenue streams and increase the productivity of their operations. The availability of open GPS data in the 1980s created whole new businesses in consumer and business GPS and mapping services that today contribute an estimated $90 billion a year in value to the US economy.[11] Other types of open data made available by governments, companies, and third parties (often combined with other data) can be used to create innovative products and services. Particular value may lie in bringing together information from disparate sectors of the economy and combining that information with proprietary data. Companies hoping to capitalize on open data opportunities can seek out and inventory sources of liquid data, develop strategies to influence others to make valuable data more available, invest in the development of tools that can transform open data into products and services, and dedicate staff to developing insights and designing products using open data.

- **Companies.** Open data can bring fresh insights into how companies operate and help management identify unnecessary variations and other barriers to productivity. Companies can choose to share proprietary data to create benchmarks that can improve overall industry performance. Using open data, such as customer discussions on social media, companies can refine product requirements and create new products and services. Companies will also need to have strategies about how, when, and under what circumstances they open their data, taking into account the potential impact the release of their data could have. Companies can choose platforms for data release, participate in the creation of an ecosystem of data users, and consider ways in which to monetize the value of liquid data. In addition, they can participate in standards setting, including for metadata, and provide input into the emerging legal frameworks governing data.

  Companies must also evaluate the potential risks posed by open data. Firms should become aware of the types of open data that might lead to reputational harm, hurt their competitive position, or disrupt their industry. Given these risks, companies should participate in the dialogues that set standards, develop legal frameworks and policies, and inform the broader public on such topics as intellectual property, privacy, and confidentiality.

  Data brokers provide data aggregation services and develop other services related to open data (for example, by integrating open data with proprietary data sources, then analyzing those data). Data brokers will have to continue monitoring the open data landscape for more data to collect and aggregate, and continue to develop innovative products and services, as well as participate in policy dialogues.

11 Philip Yam, "How to kick-start innovation with free data," *Scientific American*, March 23, 2013.

- **Individuals.** Consumers can capture a large portion of the value potential of using open data. Individuals can seek out applications that use open data and provide feedback to improve these tools. They can also take advantage of MyData to guide their behavior and ensure that data are accurate. Constituencies can also be built to press for release of more open data and creation of more tools from both government and other enterprises.

  While open data has numerous potential benefits, consumers can help to safeguard their interests by monitoring privacy policies and practices to ensure that the data are not used in such ways that lead to social or financial harm, or information that they simply prefer to keep confidential. Citizens can work with government to guide policy around the collection and use of data.

- **Non-profits and NGOs.** Open information on the location of resources—schools, hospitals, roads—and on the quality of health, educational, and economic systems can be used to identify areas of greatest need and calculate the additional resources required. Non-governmental organizations (NGOs) and other non-profit organizations can incorporate this type of information when developing a strategy, and in day-to-day operations. The first step to this will be understanding the data that are available and identifying key gaps. Lobbying governments and private organizations to release information where gaps exist can lead to critical breakthroughs.

  NGOs and non-profits can also organize volunteers with data-related skills who can collaborate to create useful tools. The OpenStreetMap database that was used in Haiti, for example, was produced by volunteers on several continents. NGOs can establish common platforms, set standards, and motivate volunteers. Non-profits can also help fill the talent gap: Code for America, for instance, recruits web developers, designers, and entrepreneurs for a year of service helping cities to use the Web to become more open and efficient. Non-profits also can serve as a neutral organizing force to align international data standards, such as standardizing formats for metadata (data about data) and other data elements.

  While open data can increase the impact of outreach efforts, NGOs and other non-profits must also be aware of the risks and be careful that the data they collect not inadvertently identify or reveal negative information about groups that they are trying to help. Steps should be taken to ensure that privacy is maintained, especially as numerous de-identified data sets are combined, making it more difficult to mask identify.

□ □ □

Our research shows that open data efforts could help to unlock more than $3 trillion in value every year in seven domains of the global economy. The benefits of open data can be self-reinforcing: as individuals perceive benefits from the use of open data, they will help to improve the accuracy and detail of information available, thus increasing the value of the data and the benefits that they can receive. However, this cycle can gather momentum only if private industry and public agencies cultivate a vibrant open data ecosystem and create data policies that provide adequate protection for all stakeholders. Companies will need to put in place the technologies and talent to collect and analyze the data. Individuals—as consumers and citizens—will need to be vigilant and savvy providers and users of open data.

1. Open data in education

Today, more than one billion students are enrolled in schools around the world, served by more than 58 million teachers. Annual spending on K–12 and post-secondary education (the scope of this report) exceeds $4 trillion worldwide.[12] With so many resources dedicated to public education, there are substantial opportunities to increase the efficiency and effectiveness of current systems. By standardizing and sharing data that already exist, the effectiveness of education around the world could be greatly improved.

The application of open data is already changing the ways in which teachers teach, students learn, administrators make funding and spending decisions, and graduates find the right employers. We have identified five levers that can enable potential value from using open data in education: improved instruction, better matching of students to programs, matching students to employment, transparent education financing, and more efficient system administration.

- **Improved instruction.** Instruction can be improved by using data on student performance and learning styles to design and personalize lessons suited to individual skills and learning styles. The lesson plans are continually refined by analyzing large amounts of individual student data shared across systems and institutions, guiding educators to the most effective teaching methods.

- **Matching students to programs.** Open data can also be used to help parents and students identify the best fit in terms of school or program. By analyzing publicly available data on school performance (for example, educational outcomes and program offerings), they can select educational opportunities that match the student's interests, abilities, and needs.

- **Matching students to employment.** Employers and job candidates can find one another more efficiently with tools that use open data to match the skills employers need with the skills candidates possess. Firms can analyze data on their employees to determine skills most necessary to succeed at the job, then search for candidates with proven credentials, such as Open Badges, in those specific areas. Similarly, students can locate and apply for positions that fit their skills and interests but that may previously have gone unnoticed.

- **Transparent education financing.** Open data allows for more transparent education financing, which also can expand the range of education options for students. By using financial aid tools that analyze open data on the true cost of education at various institutions as well as different financing options, parents and students can better understand trade-offs and choose programs based on the actual expected cost of education. This allows students to enroll in programs they may have previously considered out of reach, and it helps

12 In defining the scope of the effort, we include K–12 and higher education that is delivered through public and private systems in the United States, the European Union, Asia-Pacific, and across emerging markets.

students who are deciding whether they can afford higher education to find suitable programs.

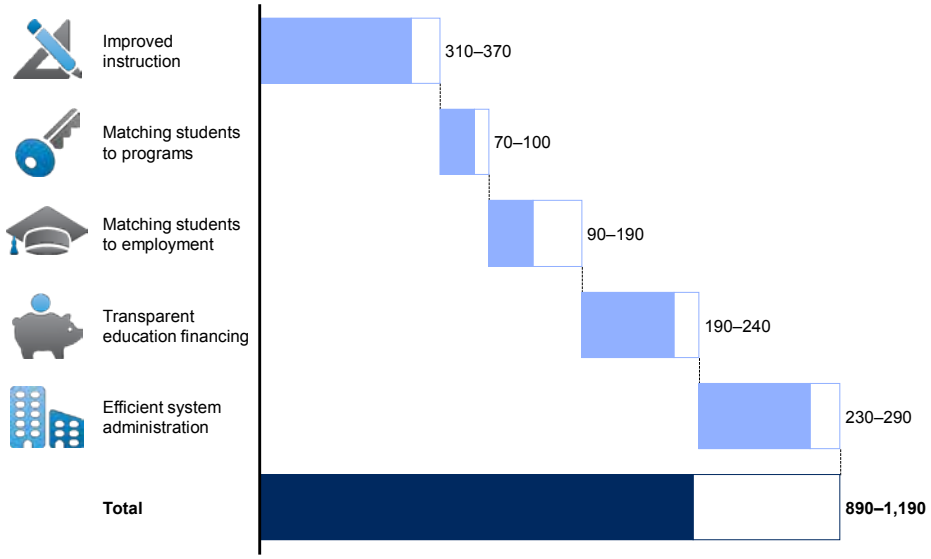
- **Efficient system administration.** By using open data to expose variation in prices paid for supplies and services and benchmarking between districts, school systems can greatly improve the efficiency of school procurement. And, by using open data to identify locations for schools based on current and future population shifts, administrators can allocate education resources in a way that ensures high utilization and provides an intelligent plan for future school investments.

In combination, these five levers can deliver $890 billion to $1.2 trillion in additional annual value (Exhibit 1). This value comes largely in the form of higher future expected earnings, but opportunities to save costs on existing operations are significant as well.

Exhibit 1

Five levers can lead to $900 billion to $1.2 trillion per year in value in education

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

Realizing the full benefits of open data in education will depend on many factors. Stakeholders will need to resolve privacy concerns by ensuring that data can be shared across the education system to help identify areas ripe for improvement, while still protecting the identity of individual students and teachers. Standardized measures will be needed for analyzing student, teacher, and school performance, and for identifying best practices. These measures should track observable metrics such as student outcomes and relevant information provided by the student, such as goals and preferences. Together, these data sets can help educators learn about learning—using data to see which approaches work best for students and refining these techniques based on ongoing monitoring.

There are many sources of education data including academic records, student preferences and goals, student and family financial data, and student-controlled MyData (see Box 3, "Sources of open data in education"). These pieces of information can be analyzed to tailor personalized learning plans, select programs that fit students well, and identify financial need and funding opportunities. Creating ways to integrate student-controlled MyData into open data platforms will be critical to unlocking the full value potential in education, and systems will be needed to take advantage of these data.

Box 3. Sources of open data in education

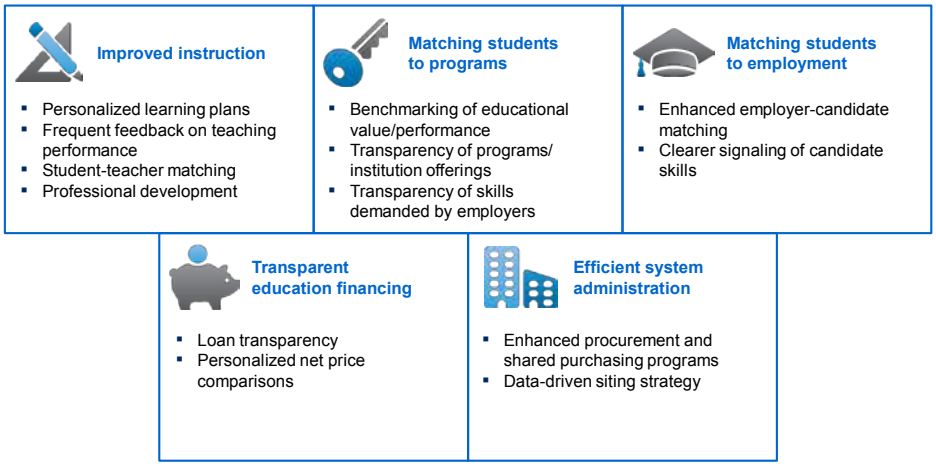
- **Educational institutions.** Schools and other providers of educational services collect data on student and teacher performance, program offerings, budgets, special programs (accelerated learning, remedial services, or after-school offerings), procurement, utilization rates, and learning standards.
- **Education agencies.** Local and national agencies are sources of information on learning goals, standards (admission and graduation requirements), system-wide testing results, funding, attainment (graduation rates), and data about the education system (number of students, teachers, schools).
- **Other government agencies.** Other relevant data collected by governments include population and economic reports, such as GDP and unemployment rates, and labor market data, such as statistics and forecasts about demand in different occupations.
- **Application developers and data services.** A growing range of ratings, learning platforms, and other education-related information services and applications are being developed. These include net price calculators, adaptive learning platforms, and school ranking systems. In addition, in the course of providing their services, education suppliers may collect data, as permitted, on student and teacher performance, school purchasing activity, and the use of support services that would be helpful to schools.
- **Individuals.** Students, families, and teachers all have relevant data about their performance and their learning and teaching styles. This includes data on their aptitudes, interests, and goals. By sharing data on personal preferences, stakeholders can learn what practices have been shown to best support their unique goals. Revealing personal preference data to neutral brokers allows for learning plans, student fit, and employment opportunities to be customized for the individual.

HOW FIVE LEVERS CREATE VALUE IN IN EDUCATION

We identify five levers for creating value with open data. These range from using open data metrics to help create customized learning plans for students to using open data to bring transparency to college costs (Exhibit 2).

Exhibit 2

How five open data levers create value in education



SOURCE: McKinsey Global Institute analysis

Improved instruction

The classroom is where many of the most valuable applications of open data will evolve. Improved instruction can be enabled by developing personalized learning plans for students, frequent feedback on teacher performance, and targeted professional development programs for educators based on continuous feedback. Frequent measurements of student performance and teaching practices can be combined with information on learning requirements, demographics, and learning preferences to identify areas for improvement. This analysis can help students, parents, and teachers improve their performance, which can lead directly to higher graduation rates, greater educational attainment, and increased lifetime earnings. These analyses can be improved when individual student and teacher data are combined with data from multiple sources (e.g., when more liquid data is brought to bear).

The data-driven approach to improving teaching and learning allows for targeted, rapid responses to issues that arise. By using insights gained from frequent diagnoses, problems can be identified and corrected before student performance suffers. Instructors and online teaching systems are able to adapt lessons and plot student learning plans based on information gleaned from a broad and comprehensive database of student profiles, behaviors, and interactions with learning materials. Sharing such data across institutions and school districts—making the data more liquid—allows for more robust analysis. For example, educators can monitor student attainment frequently and at a very granular level (measuring mastery of a particular concept, for example) and track how individual student performance correlates to classroom activities (e.g., lecture versus exercises or group work). These approaches can also be used to increase motivation, change mindsets, and adjust learning strategies to improve student achievement (see Box 4, “Tech-enabled personalized learning”).

Box 4. Tech-enabled personalized learning

Adaptive technology is a growing trend, with some 40 companies, such as Knewton and DreamBox Learning, marketing adaptive learning systems to schools in North America, Europe, and Asia.[1] Adaptive learning attempts to address the limitations of conventional classroom teaching by capturing information about what each student knows and crafting custom lesson plans based on individual knowledge and progress. Instead of marching the class through the same lesson (and often covering the same material repeatedly until most of the students get it), adaptive learning claims to deliver the right content, at the right time, in the best way for each student.

Adaptive learning systems use online lessons to teach a specific lesson or concept and monitor student progress in real time by recording indicators such as the number of times the student pauses a lecture, the time needed to answer a question, the number of times a question was attempted, and a student's reliance on "hints." This results in a rich set of data that can deliver insights on which concepts a student has mastered, where more work is needed, and what specific interventions might support greater mastery of concepts. Teachers can use these insights to intervene and further personalize student interactions.

Adaptive learning platforms have been used at all levels of education and training. At Arizona State University, an adaptive learning program helps students who are struggling with remedial math. Student pass rates have improved from 64 percent to 75 percent, and dropout rates are down by 7 percent, ASU reports.[2] The United States military uses adaptive learning software to train submarine technicians and finds that new recruits using the system learn more and perform tasks better than technicians who have been on the job for seven years or more.[3]

Implementing adaptive learning faces several challenges. As teaching becomes more adaptive, teachers and professors might need to adapt to doing less lecturing and more tutoring and coaching. Lessons delivered by computer can be broken down into multiple personalized modules, instead of set class periods. Students may spend 15 minutes reading new material through an adaptive program that measures how they are interacting with the content, then work in groups or on personalized lessons based on their earlier interactions. This may be followed by a tailored quiz and a homework assignment to reinforce specific points. Such modifications rely on comparing the progress of a particular student against that of others with similar learning needs, backgrounds, and support. Since students can progress at their own rates, the traditional "semester" and "grade" model may also need to evolve. This has implications both for how K–12 classrooms are organized and for how colleges account for credits and graduation requirements.

1 Education Growth Advisors, *Learning to adapt: Understanding the adaptive learning supplier landscape*, Bill and Melinda Gates Foundation, 2013.

2 Data on Knewton performance in remedial math programs at Arizona State University from www.annualreport.asu.edu.

3 Office of Naval Research, "Submarine piloting and navigation training" exhibit fact sheet, www.navair.navy.mil.

Opening and sharing student data across the educational system help develop more accurate and robust insights into what works and what works best for particular types of students. If these data are obtained from a broad base of students and insights from the data are made available to educators, the findings can help teachers identify more clearly which practices are working best in their classrooms before the end of the term, so they can use the most effective ones and de-emphasize others. Similar information can also help to identify the most valuable professional development opportunities for teachers. Teacher performance data can also help match teachers with coaches who can help them grow as professionals.

To implement improved instruction, teachers and schools need open data platforms that provide access to machine-readable student and teacher performance data, lesson plans, and demographic data. Concerns about privacy and misuse of data must be addressed, and safeguards must be in place in order

to enable these benefits. There is significant potential for outsized gains in student performance from personalized learning approaches, which have been proven effective in other settings and now can be brought into K-12 schools.

We estimate the potential value from improved instruction to be $310 billion to $370 billion per year worldwide, largely through increased lifetime earnings. This estimate includes the potential benefit of adaptive learning programs over large populations of students.

Matching students to programs

Giving students and families open data about school performance and program offerings can help them identify the best educational opportunities, especially in the selection of post-secondary schools. Side-by-side comparisons of graduation rates, student test results, expected incomes of graduates, areas of specialization, and program choices can give students a clear picture of which institutions would be best for them. Many schools in the United States publish school and program descriptions, along with data on school location, demographics, performance, graduation and completion rates, and ranking. This information allows parents and students to compare alternatives. Recommendation engines could be built on these open data to suggest schools and programs for individual students based on demographics, stated preferences, and achievement data.

College Scorecard, a US government initiative, aims to consolidate data about colleges and costs on a national basis, to help high school students and families select colleges and other post-secondary educational institutions. College Scorecard combines school performance data with information about future lifetime earnings for graduates. Used in combination with other data and tools such as net price calculators, College Scorecard data can help students and families identify the trade-offs among different schools in terms of cost, quality, and opportunity.[13]

School choice is already widespread in many countries in the Organisation for Economic Co-operation and Development (OECD) group of developed economies and is becoming more common in US K–12 school systems, as are open data resources to help families determine which schools they would prefer.[14] In Boston, where students and parents can list their top eight public school choices, an open data app for smartphones lets them see which schools are the best performers and offer the specialized programs that they want. Boston's revised choice program was scheduled to go live in the 2013–14 school year.[15]

In North Carolina, the Charlotte-Mecklenburg Schools runs a school choice program in which the district shares detailed data on schools, program offerings, location, standardized test scores, and rankings. Students and parents are then given the opportunity to rank their school preferences. Since the school information has been made public, enrollment in the district's higher performing schools has risen by 6 to 7 percent. It is important to note that while school choice in K–12 may help students find institutions that can better serve their needs, many districts cannot offer every student a slot in a high-performing

13 Future earnings by major are not yet available from College Scorecard.

14 "How does school choice work in other countries?" Friedman Foundation for Educational Choice, August 2013.

15 Boston School Choice, www.bostonschoolchoice.org.

school. For example, spaces in top schools are so limited in Washington, DC, Seattle, New York, and Boston that these cities run annual lotteries to determine which qualified students can fill available slots. Open data can also help families identify alternatives and put pressure on systems to provide more options for high-performing students.

Overall, we find that using open data to help match students to schools and programs could help enable $70 billion to $100 billion in annual value. These estimates include the higher lifetime earnings that can be expected for students who switch to higher performing K–12 schools, using open data to make their choices.[16] This estimate also includes improved lifetime earnings for students who rely on open data to select better post-secondary programs and the increased productivity across the economy that would result from students making the educational choices that prepare them for jobs that will be in demand.[17]

Matching students to employment

The inability of employers to find qualified job candidates and the failure of graduates to land jobs exact significant economic and social costs.[18] Open data can help reduce this labor market problem by providing students with clear information about what skills and qualifications they should acquire to meet employer needs. Open data can also help job applicants signal to employers what skills they possess. This can help prevent mismatching in the labor market and, potentially, improve employment prospects for candidates. Today, school reputation is often used by many employers as a proxy for a candidate's skill level, which can lead to mismatching and costly turnover. Open, verifiable data on the actual skills a candidate possesses could significantly improve the job search process and also decrease turnover, since candidates are more likely to succeed in the roles that they take on.

Open data records that communicate student skills can help candidates signal what their actual credentials are. Mozilla's Open Badges platform allows candidates to collect badges that have openly verifiable requirements, such as performing well on a test, to demonstrate proficiency in a certain programming language, for example. This badge information can be used in several ways. First, employers and large labor agencies can analyze the badges claimed by their own employees to help determine the skills and qualities associated with successful job performance. By understanding what it takes to succeed, employers can better target future recruiting efforts. Employers may scour credentialing databases for candidates who can prove that they have the most useful skills for a given role.

Candidates can also find out about roles that they had not considered but that require skills they possess. This can reduce search costs and may also help limit one cause of turnover, since better matching should result in more employees

16 Improved lifetime earnings from attending better schools was limited to 70 percent of students since many districts do not have enough seats available to meet demand for higher performing schools; see Eric Hanushek, "Valuing teachers: How much is a good teacher worth?" *Education Next*, volume 11, number 3, Summer 2011.

17 We include benefits associated with preparing for more in-demand jobs because it will require students to seek particular educational opportunities. Reductions in friction of finding employment are captured in the "matching students to employment" lever, although previous skill acquisition certainly plays a role in that lever as well.

18 See *Education to employment: Designing a system that works*, McKinsey & Company, December 2012.

who are happy and successful in their work. It would be especially helpful if this sort of matching worked both ways, with employers searching for skills on common platforms to identify potential candidates, further reducing the cost of candidate lead generation. ACT WorkKeys, a US website, lets students and job candidates demonstrate their skills through an online testing and assessment process. Since 1990, it has matched its assessment scores to more than 18,000 types of jobs in an online occupational database offered to employers.[19] Candidates also learn which jobs they are qualified for and what kind of further training and education may be required to get the jobs they want.

Addressing potential skill gaps and labor mismatching present a major opportunity to create value with open data. Annual turnover rates are well into the double digits globally, and in some countries staff turnover averages 25 percent per year.[20] By providing a network where many candidates can show off their abilities, open data can help reduce friction in hiring. We estimate that the potential value from improved matching of candidates and employment is $90 billion to $190 billion per year, based on reduced turnover due to better matching. Turnover-related costs include more money spent on recruitment and training, as well as lost productivity. We consider this to be a conservative estimate, since we do not take into account the value of reduced search time.[21]

Transparent education financing

One of the murkiest areas in higher education (and in private K–12 education) is the process of finding student aid or securing loans, scholarships, and grants to help cover tuition costs and school fees. To realize the value from transparent financing, students and parents need to be able to understand how their individual financial positions fit with aid programs and have a picture of the true cost of education (accounting for any scholarships, loans, or grants). Financial aid clearinghouses that help match funding sources to individual student needs could help reduce the growing burden of student debt, increase the number of students who are able to pursue a better education, and help reduce the number of dropouts due to financial reasons.

Open platforms for scholarship and loan data could help prospective students identify sources of funding that may otherwise go unused. Similarly, loan engines that are fueled by open data, which would be similar to the personal finance engines that help consumers compare mortgages and credit cards, could help students identify the best loans for them, based on criteria such as maximum available loan amount, interest rates, payment options, and penalties. Students often do not maximize their low-cost funding opportunities because they lack sufficient knowledge about what is available. An advanced comparison engine could provide a view of all types of loans, limits, and terms, and also help families understand whether they will qualify for favorable rates, based on their credit history, expected earnings, and other information.

One of the biggest questions for students and families—which open data is helping to answer—is the actual cost that they are likely to pay. In the

19 www.act.org.

20 Shreya Biswas, “Attrition in India to top world charts in 2013; one in four employees to change jobs,” *The Economic Times*, June 7, 2013.

21 Systematic search and targeting of candidates by employers is still in a very early stage of use, so we do not size it here.

United States, the difference between stated tuition and the net price—what a student is responsible for paying after financial aid is subtracted—can be $40,000 a year or more. Since 2011, the United States has required all schools to provide a net price calculator to let students estimate their personal cost of attending for the first year of college, based on an estimate of aid they are likely to receive. Students enter their Free Application for Federal Student Aid (FAFSA) information, which includes family income and other factors, to come up with an estimate of aid and net costs.

A startup called College Abacus now aggregates net price calculator data from more than 2,500 institutions—from community colleges to Ivy League schools—to allow students to estimate the cost of education and see the trade-offs before they apply. Parents and students become better informed by having simple tools that are based on open data and are easy to use. By helping students locate funding for a program that was assumed to be too expensive based on full tuition prices, open data can help a student attend the school that has the most appropriate programs and provides a path to higher earning potential. Exposing true college costs can help today's students save money, increase the number of students who go on to post-secondary education, and eventually influence the quality and the value of schools. We estimate that the potential value of transparent financing is $190 billion to $240 billion per year.

Efficient system administration

Open data can be used to reduce school operating costs, by exposing variation in prices between vendors and across school systems, enabling shared savings programs, improving supply/demand management of goods and services, and determining the optimal school location strategies as populations shift.

While schools can reap many of the benefits of enhanced procurement using their own data, sharing their information across schools and districts can lead to even better results through benchmarked price negotiations and providing better visibility into how districts have reduced the use of various goods and services. Today many K–12 schools report procurement costs to school boards and governments, but often in insufficient detail to make meaningful cost comparisons. By sharing data on school spending, districts can benchmark their performance against a peer set. This allows for overpayments to be identified and targeted for cost reduction.

Improved purchasing of broadband services is an example of how benchmarking can reduce operating costs. Many schools pay flat, uncompetitive rates for broadband services, without soliciting competitive bids. Estimates in the United Kingdom suggest that schools pay three to four times what private firms with similar capacity requirements pay for broadband service.[22] By comparing the price of broadband service offered by local suppliers, a school district in northern California was able to reduce broadband costs by 30 percent.[23]

Savings in procurement can be achieved by using open data to consolidate, standardize, and compare the types of products purchased based on quality, price, and other value metrics derived from open data on budgets, contracts, and transactions. By allowing administrators to analyze past purchases and

22 "Procurement gap," Education Superhighway, www.educationsuperhighway.org.

23 "The broadband agenda: School internet … for less," RM Education, www.rm.com.

examine how peer institutions are spending their budgets, open data can also enable better forecasts for when products or services will be required. With the correct data, administrators can identify areas for standardization of goods and consolidation of purchases.

Romania has used open data to reduce the price schools pay for a wide range of supplies. Its open bid platform hosts an online auction, where vendors can bid on contracts to supply 2,000 basic goods for schools across the nation. Suppliers see information about the quantity of goods requested, current bid offers, and data on previous auctions, such as closing bids, prices, and quantities. Two years after implementation, the average price of the items offered on the platform had dropped 24 percent. Similar efforts in the United Kingdom yielded savings of 8 to 9 percent on items acquired through the open bidding system. These efforts not only reduce procurement costs, but also allow vendors to better forecast demand and also identify opportunities for new products or services.

According to the Illinois Association of School Boards, reverse and open bid auctions helped save 12 percent on commodity purchases such as copy paper.[24] Within universities, there are additional opportunities to use data to cut costs, for example by centralizing purchasing of lab supplies that are used in both nursing schools and in biology and chemistry departments. Several groups of universities have begun to share procurement data for goods and services.

Open data can also help inform decisions for rationalizing the number and location of schools within an area, which can lead to significant savings. Running schools that are operating below capacity in areas where the school-age population is declining can be expensive, especially when schools are housed in older buildings that are not energy efficient. Using open data about demographic and employment shifts within the community and area, school officials can predict future enrollment and estimate operating costs (for example, heating and electricity bills, maintenance costs) as well as the potential impact of school closings, such as disruption to families who will need to switch schools.

The Edmonton School District in Canada, for example, is using this process to consolidate its 199 schools over the next five years, with a goal of increasing utilization and reducing operating costs.[25] Planning school footprints based on current and future demographic needs of a community requires access to population data, school utilization rates, and expected economic growth. It is important to note that these school closures do not come without risks or their own additional costs.[26] However, open data sources can help administrators intelligently plan their future school footprints.

Overall, more efficient system administration has the potential to save districts and institutions $230 billion to $290 billion per year, approximately 8 to 10 percent of public education spending. The majority of these savings come from enhanced procurement. Smarter school footprint management also makes a significant impact. Consolidation and smart-growth plans can help all kinds of districts—rural

24 Jim Headlee, “Reverse auctions revolutionize purchasing,” *Illinois School Board Journal*, May/June 2010.

25 Ten-Year Facilities Plan, 2009-2018, Edmonton Public Schools, www.epsb.ca.

26 Many parents are concerned about children’s safety and inconvenience when children are re-assigned to schools that are farther from home, and the potential loss of a local school can serve as a unifying force in the community.

and urban, large and small—to save money on facilities and find more funds for teaching.[27]

BARRIERS AND RISKS

Open data clearly presents significant opportunities to create value by raising the quality of education, reducing the cost of administration, and improving the skills of the labor force. Students and parents can use open data to find the best K–12 schools, then use open data to select the best post-secondary program (and the best way to pay for it), and eventually use open data to find the employer that offers the most rewarding job opportunities, based on an individual candidate's skills. The result can be rising skill levels and better matching in the labor market, which can raise productivity across the economy.

However, few of these benefits can be realized without addressing important obstacles, including resistant mindsets, the relatively primitive state of open data in education, and significant gaps in technology, funding, and technical capabilities. Without support from the key stakeholders—students, parents, teachers, administrators, and policy makers—open data efforts will not go far. Nor can open data efforts succeed without investment in technology and new capabilities in the classroom and in educational administration.

Parents, students, and teachers—and society at large—have legitimate concerns about how educational data are gathered and used. Will a student's longitudinal performance data, intended for educational professionals and researchers to improve instruction, become public or be used by private companies to earn a profit? Could students—particularly those with relatively low performance—be denied educational and employment opportunities if data become public? Teachers and administrators are concerned that data about their performance—or of their students—may be used out of context. In some cases, these concerns have already slowed or halted implementation of data-driven approaches.

For open data to advance in education, there also must be agreement on what to measure, how to measure it, how to share it, and how to use it. Currently, for example, there is significant debate over which student and teacher metrics capture the effectiveness and impact of teaching methods. Even when there is agreement on what data should be collected, a lack of standards for how the information is generated and formatted and how terms are defined (e.g., freshman or ninth grader) make apples-to-apples comparisons difficult across schools and even within schools. Furthermore, summary data are often presented to the public in confusing ways that make it difficult for families to use the information correctly. Many reports are difficult to read and do not provide the proper context; in some places, for example, scores of special-needs students can skew average school scores.

Data sharing across schools is also impeded by a lack of clarity over ownership and privacy. For example, private vendors seek to protect competitive intelligence gained through administering tests, while schools seek to keep data to themselves to protect privacy and use it to improve their performance internally. In some cases, administrators wish to prevent disclosure of data about poor performance. However, insights into the most effective teaching and learning

27 Some 40 percent of US school districts have not gone through consolidation efforts in the past 20 years; see Thomas Snyder, *120 years of American education: A statistical portrait*, National Center for Education Statistics, US Department of Education. 1993.

approaches can be established only by analyzing large samples to capture a range of experiences and outcomes, making sharing of data critically important.

Finally, open data cannot really take root in education until the proper infrastructure, technology, and capabilities are in place in schools. Today, many schools still do not have the technology required for data-driven approaches. While 97 percent of US public and charter schools have basic access to the Internet, many schools have the equivalent of residential connections, which is not enough to support Web-based adaptive learning. Educational institutions often lack the skills necessary to incorporate data into their everyday workflows, such as data analytics experts, data visualization specialists, and managers who can inculcate a data-based culture. Schools also lack leaders with the skills to administer data-driven learning systems and to capture and report data from their schools.

IMPLICATIONS FOR STAKEHOLDERS

Key stakeholders will need to consider actions necessary to fully realize the value from open data value and overcome the obstacles and risks described above.

- **Students and families** will need to take greater ownership over each student's educational path. They can use performance dashboards, reports, apps, and other sources of information about their own performance and schools to make more informed decisions about the student's learning or use this information to lobby for specific improvements at a child's school. Students and families can also participate in the design of open data platforms, providing feedback and suggestions about the tools and reports they see or wish to see. To get the most out of open data, families and students must take on roles as advocates, demanding more evidence-based instruction and the funding needed to support it.

- **Teachers** must be on board; their participation and support are critical. Teachers and professors can become excellent sources of input for how schools and third-party vendors can use data to improve teaching technology and help the data analysts refine metrics used for student and teacher assessments. Teachers can also invest time in becoming more familiar with the technology available and the metrics used to report student and teacher performance. The benefits of improved instruction will be realized only if open data-enabled tools and techniques are embraced in the classroom. For example, at Arizona State University, pass rates in remedial math were as low as 33 percent in classes where professors did not encourage students to use the adaptive learning materials, but ranged as high as 100 percent in classes where professors embraced the program.

- **Administrators** can advance data-driven approaches in their institutions and help build trust and confidence in the new techniques with teachers, parents, and their communities. To start, they should be clear about goals and transparent about program design, implementation, and outcomes. By sharing the purpose of an open data tool and building feedback from users into the design, administrators can alleviate concerns and confusion. When their institutions or districts commit to open data, administrators will need to secure the financing, technology, and talent required for successful implementation. They can seek infrastructure assistance through vendor programs and public-private partnerships and resources such as School of Data, a non-profit that helps organizations acquire data skills.

- **Application developers and suppliers** can play an important role in open data into education and matching students to employment, through tools such as MyEdu or After College. They will have to work closely with teachers, administrators, parents, and students to prove that their interests are aligned, by making clear the value of open data for all stakeholders and being transparent about exactly how data will be used. Application developers can build databases that integrate open data records. They can develop the infrastructure required to measure student performance and create adaptive learning tools. Third-party vendors can help to formulate and adhere to open data standards so that a large and robust base of knowledge can be built that draws data from classrooms and institutions across the globe. Vendors can also build support by soliciting teacher, parent, and student input into improving their products. To make their offerings more valuable, vendors will need to invest in intuitive interfaces and dashboards and in highly secure technology platforms that adhere to or exceed guidelines on data security of student and teacher information.

- **Policy makers and regulators** can lead the discussion about the role of data in the classroom and can educate voters on the positive impacts of data-driven solutions in education. They can also create policy that supports and encourages the use of data. Successful advocacy on the part of policy makers can lead to additional funding to foster the adoption of open data-enabled programs in public education. Laws and regulations governing the use of open data will need to balance privacy and property protection with the need for gathering data in sufficient amounts and with sufficient granularity to make open data useful. Lawmakers and policy makers can help determine standards for reporting, formatting, and distributing data across jurisdictions. They can also lead the public debate about the trade-offs between privacy and the sharing granular data and establish guidelines to govern the use of personal data. In the United States, FERPA regulations (based on the Family Educational Rights and Privacy Act) already establish rules about student and family information. Even more stringent, the European Union's Data Protection Directive provides expansive guidelines on what personal data can and cannot be shared or made public. Finally, lawmakers will be instrumental in funding infrastructure investments. The US E-rate program has been instrumental in funding the wiring of 97 percent of schools in the United States, but other investments will be needed to upgrade this infrastructure to support online learning platforms in the classroom.

- **Employers** can benefit from open data in education by finding better workers and reducing the costs of recruitment. To realize these benefits, they will need to partner with schools to help design curricula and programs. Employers will need to clearly communicate the skills necessary for a given position and support the development of tools and standards to match candidates to jobs (as in using badges to pre-screen candidates). A top priority now should be to determine which metrics are best for documenting that students have the correct knowledge and skills. As users of educational open data, employers can be a useful source of feedback to guide improvements in reporting. Employers can also help schools understand the effectiveness of their instruction and how well they prepare students for work, as well as helping schools to identify gaps in learning and define emerging skill and training needs.

2. Open data in transportation

Transportation services provide critical links in the global economy. Every year, nearly nine billion tons of cargo—about 90 percent of international trade volumes—are transported by ship. Some 90 million flights a year move billions of passengers and more than $3 trillion in goods. Every year, railroads make 40 billion passenger trips and carry 30 million railcar loads of cargo. More than 11 million miles of paved road enable one billion cars and trucks to move passengers and cargo across the globe. These four modes (marine shipping, air, rail, and automobile) provide $4.5 trillion in transportation services per year.[28]

Massive opportunities exist for efficiency improvements in each of these modes of transportation, which could save billions of dollars in direct costs and add many billions of dollars more of value to the global economy by reducing delays. In the United States, the Federal Aviation Administration estimates that flight delays waste $4 billion in fuel each year. In Europe, 20 percent of trucking trips carry no cargo. For commuters in the United States, travel time to work can vary as much as 45 percent on any given day, wasting up to 50 hours a year per commuter in delays and "buffer time" (the extra time required to ensure smooth transit).

Open data has already played an important role in improving transportation and we believe that it has significant potential to play a role to in addressing the inefficiencies that persist. Many of the improvements we have seen in supply-chain management, logistics, mapping, and route planning over the past two decades were enabled by opening government data from GPS satellites. Transportation customers (defined here as individual travelers as well as organizations and parties that ship cargo) benefit from improved tracking of people and freight. Data streams and data sets on bus and train arrivals, ship positions, and aircraft utilization rates are being made open and allow individuals, firms, and government to provide and use services in new ways.

28 We consider the impact of open data on marine shipping, and air, rail, and automotive transportation, as well as related infrastructure. Effects on adjacent industries such as tourism and hospitality, and transportation of fuel, water, and electricity are beyond the scope of our research.

There are three major levers for unlocking value with open data in transportation: improved infrastructure planning and management; optimized fleet investment and management; and better-informed customer decision making.

- **Improved infrastructure planning and management.** Open data on passenger flows and door-to-door travel times allows network operators (including municipal transit systems) to improve capacity and throughput. Open data about cargo destinations and system bottlenecks can be used in daily operations as well as in planning for new infrastructure, such as rail line extensions or port expansions.

- **Optimized fleet investment and management.** Real-time open data about vehicle location and condition and benchmarking of vehicle cost and maintenance information can help operators purchase, deploy, and maintain fleets more efficiently.

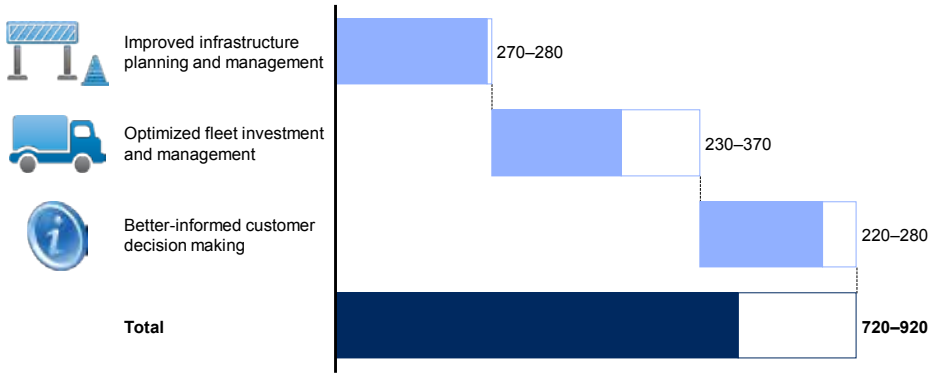
- **Better-informed customer decision making.** Detailed open data about costs, reliability, environmental impact, and other factors can allow customers to make better decisions about which mode of travel to use and when.

Based on our analysis, we estimate the global potential economic value that could be unlocked through these open data levers in transportation to be $720 billion to $920 billion per year (Exhibit 3). Optimized fleet operations (fuel savings, more effective maintenance, higher utilization) could enable as much as $370 billion a year in value. Improved infrastructure planning and management and improved consumer decision making can each lead to value of as much as $280 billion per year.

Exhibit 3

Three levers can lead to $720 billion to $920 billion per year of value in transportation

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

With billions of passenger trips, tens of millions of rail freight cars, and 18 million seagoing cargo containers, the transportation sector generates enormous amounts of data every year. These data are captured by travelers, government, network operators, and third-party data vendors (see Box 5, "Sources of open data in transportation").

Box 5. Sources of open data in transportation

- **Government.** Government agencies are the primary sources of open data relevant to transportation planning and operations. This ranges from geospatial data (e.g., maps, GPS signals) to weather data (including records of natural disasters), to industry profiles and growth projections. Governments also collect detailed demographic data through census and other research that can be used in infrastructure and route planning. Information about utility rates and vehicle reliability, such as recall statistics, are also published by government agencies.

- **Transportation operators.** Shipping companies and operators of bus, rail, and airline systems all generate a great deal of data that, if shared, can help manage daily operations of connected systems, such as airports and airlines, as well as investments in equipment and infrastructure. This information includes figures on vehicle safety and performance, inspection and maintenance data, schedules and fares, manifests, and on-time performance. Public transportation companies and agencies can provide real-time vehicle location data that can be used in consumer applications, such as delay alerts.

- **Individuals.** As they travel, consumers generate data about their use. For example, electronic fare systems on commuter bus and train systems capture origin and destination, time of day, duration of every trip, and frequency data. Toll-collection systems can provide similar data about automobile trips. Data can also come through opt-in mechanisms, in which travelers choose to reveal preferred modes of travel, willingness to pay, and the purpose of their trips. Some mobile phone applications report the speed of travel of individuals while they drive, which can then be aggregated to reveal the speed of traffic on specific roads. An application developed by the city of Boston called Street Bump uses a smartphone's accelerometer to report when a driver has hit a pothole.[1]

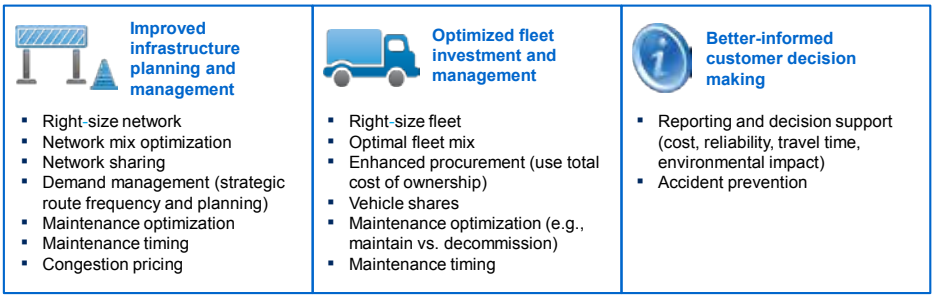
- **Third-party data providers.** Starting with publicly available data and input from customers (e.g., airlines), third-party specialists can provide data on a wide range of transportation behaviors as well as benchmarking statistics. For example, they can make comparisons among modes of transportation, based on costs, price, emissions, and other factors, as well as comparisons between operators in each mode. Third-party sources can also provide information about events and disruptions that can slow transportation services.

1 Street Bump uses data from "spot reporters," drivers who choose to report accelerometer changes recorded by their smartphones, to pinpoint roads in need of repair. To date, more than 24,000 reports have been made (http://streetbump.org).

HOW THREE LEVERS CREATE VALUE IN TRANSPORTATION

Each value lever can be applied in multiple ways. We identify several ways in which governments can improve the efficiency of transportation infrastructure and control costs, such as by using open data for pricing schemes that relieve congestion, and multiple ways for fleet operators to optimize purchasing and operations. Shippers, travelers, and commuters are likely to gain access to more real-time information on which to base decisions about modes of travel, timing of trips, and other transportation choices (Exhibit 4).

Exhibit 4
How three levers create value in transportation



SOURCE: McKinsey Global Institute analysis

Improved infrastructure planning and management

Open data can improve the processes of investing in and operating transportation infrastructure by enabling better demand forecasts, using data-driven approaches to prioritize network improvements based on estimated returns on investment, and improving the ways that modes of transportation interoperate.

Open data already has been used to improve the design of transportation networks. For example, when Moscow's transit authority was modernizing its public transit system in 2012, it depended on open or shared data to determine where commuters lived and where they worked in the Russian capital. Officials used mobile phone location data along with government information on the ages, professions, and home neighborhoods of workers who commuted to specific business districts. Moscow then used this information to determine if greater investment was necessary in rail networks or if other services could do a better job of meeting demand. Based on the research, the city decided not to make a costly investment in a new rail line and instead met transportation needs by redrawing 100 bus routes. This limited Moscow's upfront investment costs and ensured that services could be flexible enough to meet the needs of a shifting population. In addition to avoiding more than $1 billion in infrastructure costs, the new bus routes reduced average morning commute times by three minutes per trip, saving ten hours per of travel time for each rider every year.

In New Jersey, the mass transit authority (NJTransit) released data on passenger flows to the public in 2012. Third parties quickly analyzed ridership at different times of day and were able to pinpoint underutilized rail stops, which led to more express trains and a saving of six minutes from the average commuting time during rush hour.

Open data also can help raise capacity and throughput of transportation systems. The US Federal Aviation Administration is transitioning from a radar-based air-traffic control system to a "Next Gen" system that uses GPS data from transponders on all commercial aircraft. In addition to monitoring a plane's progress, Next Gen will record arrival and departure times, passenger and cargo capacity, and route data—all of which will be made available to pilots and airlines. These data can be used to help identify ideal routes, design fuel-efficient approach and landing patterns, and increase capacity by reducing the buffer space between planes required for safe operation. Information is shared with pilots and airlines to improve operations and planning. When combined with other publicly available data, for example weather and schedules of special events, airlines can help plan more efficient flight operations and airports can better anticipate demand.

Open data can also enable real-time congestion pricing and intelligent road applications that help relieve gridlock. Singapore, San Diego, and Seattle have experimented with dynamic pricing models to foster efficient utilization of the overall road network.[29] Drivers are provided with information on current travel times along various routes and toll pricing along these routes is adapted in real time to encourage travelers to seek alternatives such as mass transit at times of high congestion. Analytical models shift prices in response to route travel time until the network rebalances and travel times fall. Early findings suggest that dynamic pricing can help cities reduce traffic time on select routes by 20 to 30 percent.

In total, using open data in improved network investment and operations could help capture $270 billion to $280 billion in value. Cost savings from smarter network investments such as Next Gen air traffic control could be worth $30 billion, and increased productivity could add up to $200 billion to $260 billion, using programs such as congestive pricing or route optimization to move people and cargo through the system in a more efficient manner.

Optimized fleet investment and management

Transportation providers face high capital and operating costs. Open data can help them optimize the size and mix of their fleets and improve operational efficiency through such measures as using open data to fine-tune the frequency and timing of maintenance. In Washington, DC, open data helped the city understand how its vehicles were being used, which uncovered an opportunity to cut the fleet by as much as 30 percent. All vehicles are equipped with GPS transponders so they can be cataloged and tracked in real time. Data about the location, type, and condition of available vehicles were shared with city employees. As a result, underutilized vehicles were placed in service more often and the city was able to fulfill its needs with a smaller fleet.

In California and Texas, publicizing government budgets at the line-item level allowed outside auditors and citizens to identify opportunities for fleet vehicle

29 The congestion pricing schemes of London and Stockholm have garnered a great deal of media attention. However, they rely on set prices for travel in congestion districts and do not use open data for dynamic pricing. Seattle, Singapore, and San Diego have experimented with dynamic pricing models, using real-time open data on passenger flows to set prices and shift behavior to balance the system. For more on this, see *Good practice in value pricing*, US Federal Highway Administration, May 22, 2012, www.fwha.gov; and Puget Sound Regional Council, "Traffic choices study—summary report," 2008.

reductions.[30] Internally, administrators learned which government agencies and departments had more vehicles than they needed to fulfill their missions. Administrators also used open data from other public agencies to benchmark against. After California released budget data on vehicle spending, citizen advocates spotted examples of high costs and pressured state agencies to reduce fleet size by 15 percent. In Texas, the state estimated that benchmarking made possible by an open review of the state budget helped reduce fleet vehicle costs by more than $5 million, or approximately 5 percent of the annual vehicle purchasing budget.

Fleet operations also can benefit from open data. Using GPS information, freight and passenger fleet operators can assemble a picture of the current location, speed, route, size of load, and destination of every asset in their fleets and make data-driven decisions about how best to operate, by optimizing routes based on traffic, use, and maintenance needs to get the most out of their fleets. While currently this information is proprietary, sharing data about optimal maintenance routines, efficient load factors, and other operational data can lead to significant efficiency gains among operators.

Companies may have data about their own empty trucks, but sharing data on truck capacity, location, availability, and customer demand across fleets could help match empty or underutilized trucks with freight that needs transporting. This could have substantial impact, considering that in Europe, 20 percent of all truck trips carry empty loads.[31] In San Francisco, Uber, an on-demand limousine service, reviewed publicly available data on crime and found a positive correlation between crime rates and demand for its services. It was able to predict that neighborhoods with higher crime rates have greater demand for rides and shared this information with the independent car service owners affiliated with the service. By choosing routes to match predicted demand, drivers have raised their revenue and customer satisfaction scores, while reducing fuel use.[32]

On the seas, the Automatic Identification System (AIS) uses transponders to track the location of all ships larger than 300 tons. Marine shipping vessels over 300 tons are required to broadcast information about the passengers and freight they carry, current location, speed, route, size of load, and destination. Shipping firms now can track every asset in the their fleets in real time and compare performance with other vessels to make data-driven decisions about how to optimize routes and schedule maintenance to get the most out of their ships. Using location data, weather data, and real-time position of other ships, fleet owners can determine if they should slow down ship speeds to avoid congestion at a destination port or reroute around rough seas. These modifications allow shipping companies to save on fuel costs as well as limit risks. Data on ship draught can be used to estimate capacity utilization by competitors, allowing firms to infer which ports and routes have the highest demand. We estimate that such measures can save up to 5 percent of the operating costs, raise utilization rates, and improve safety, speed, and reliability.

30 CALPRIG Education Fund, “California budget transparency 2.0: Online tools for better government,” www.inthepublicinterest.org, and “Texas Comptroller Susan Combs says state has tightened its belt through transparency initiative,” www.window.state.tx.us/news, December 3, 2008.

31 McKinsey expert analysis.

32 Liane Yvkoff, “Neighborhoods with more crime have more Uber rides,” CNET.com, September 13, 2011.

We estimate that using open data to improve fleet investments and operations could save $230 billion to $370 billion in costs per year, based on using open data to help firms and transportation systems select the right type, quantity, and mix of vehicles. Improved fleet purchasing can contribute $40 billion to $90 billion in cost savings from increased utilization of existing vehicles and reduced need for additional ones. Using open, real-time data on weather, crime, and location can improve operations by revealing alternatives to existing practices, which can help reduce costs or improve performance (e.g., Marine AIS, Uber). This can contribute $140 billion to $210 billion in additional value. Traditional benchmarking efforts enabled by sharing operational performance metrics across peer organizations or by comparing publicly available metrics, can unlock an additional $50 billion to $70 billion in value.

Better-informed customer decision making

Transportation customers must make several decisions for every trip. Passengers must choose among modes of travel, schedules, features of travel (class of service, for example), and price. Freight customers have additional considerations, such as the need for refrigeration

To improve decision making, customers seek data that allow them to compare choices across modes and across various criteria for evaluation, such as cost, transport time, carbon impact, and reliability.

The addition of open data can help customers make better decisions, which, when applied across a large population, can be the source of significant economic value. Public transit systems have expanded the use of sensors that generate location data, which can then be used to transmit train and bus locations in real time. This information is used by transit agencies to manage their operations, but these data streams also have been made available to entrepreneurial developers such as CUBIC, whose NextBus app lets riders see current wait times and find out where there are delays in the system.

Open data on bus and train location and road congestion has great potential for shrinking the “reliability buffer”—the extra time that a traveler or shipper builds into a trip to account for possible delays, which can be as much as 70 percent of total trip time.[33] With accurate data on delays, vehicle location, and expected arrival time, passengers can choose the routes and modes that are most efficient for their needs and improve the utility and productivity of customers.

In addition to reducing travel time, open data on vehicle location and on-time performance can help attract new users, because of the increased transparency and predictability of the services. In Duluth, Minnesota, Google Transit installed transponders that recorded city bus locations and made the data public so riders could identify the location of a bus and decide if it offered a suitable alternative to driving. After the transponders were installed, ridership increased by 12 percent, reducing congestion on the roads and also decreasing the cost of transportation for many passengers.

Another way to reduce the cost of transportation to customers is through greater price transparency. Data aggregators can present side-by-side comparisons of costs, travel times, and other attributes of various options. Being able to compare

33 *Selecting travel reliability measures*, Texas Transportation Institute, 2003.

schedules, routes, and modes of transport based on price can help customers select the most efficient option for their individual needs at that time. Additional information can aid customer decision making, such as carbon dioxide emissions per passenger and average trip delays. According to one advocacy group, commuters in New York, Washington, DC, and Los Angeles who use these data to shift from driving personal vehicles to commuting by public transit can save as much as $2,500 per year in fuel, parking, and maintenance costs.[34]

To drive price transparency in automobile purchasing, third-party applications are collecting actual negotiated sales price of new cars in various regions. In California, an entrepreneurial company called TrueCar offers an "information and technology platform" that aggregates sale price data from a network of 6,000 automobile dealers to identify the actual sales prices for a given make and model. TrueCar uses this liquid data to help consumers negotiate better prices and is compensated by the dealers that participate in the program.

In total, improving customer decision making could help to capture $220 billion to $280 billion in value. Sources of value include additional time that customers gain from choosing optimal travel routes and modes of transport ($200 billion to $210 billion) and cost savings to customers who use additional information about products and prices to improve their purchasing decisions and negotiating positions ($20 billion).[35]

BARRIERS AND RISKS

There are obstacles to realizing the value of open data in transportation. Some pertain specifically to the levers we describe, while others are more general. Across potential uses, for example, we see that cost and privacy concerns are common barriers, as are gaps in data standards and capabilities in transportation organizations.

In fleet operations, the cost of equipping vehicles with sensors and transponders will require investments that may meet resistance. Also, some high-performing firms may be reluctant to share performance results that are needed for benchmarking best practices because they fear they could lose their competitive advantage. Firms also could be reluctant to submit their data on location, route, and capacity, which would be needed to locate available system capacity, for fear of attracting competition to their market. Freight companies also have concerns about maintaining client confidentiality and protecting information about what their customers are doing.

There are political and financial barriers as well. Infrastructure projects are expensive, long-run investments that often involve policy makers from multiple jurisdictions. Political considerations, rather than data-driven empirical decision making, may determine which projects are built. At the same time, there are legitimate concerns about providing transportation services to citizens in rural areas where the financial return on investment in infrastructure, which open data can help estimate, is of secondary importance.

34 "Rising gas prices equal big savings for public transportation riders," American Public Transport Association press release, July 19, 2013.

35 The transportation industry traditionally measures the value of passenger time at the rate of half of the local wage. We have used this same rate to calculate the value of time savings in the system.

One of the biggest barriers to using open data in transportation is the extent and quality of the available data. For consumers hoping to make data-driven travel decisions, there is often insufficient information to understand the complete set of options and trade-offs associated with different modes of transport or different routes (for example, airline data are reported as time in the air, not gate-to-gate time). Data are also not aggregated in ways that make comparisons easy. For the full benefits of open data to be realized, passengers and freight shippers will require easy-to-understand, meaningful metrics to make optimal transport decisions. Significant effort will need to be put into identifying the most meaningful metrics to report to consumers, as well as the best communication tools and methods.

Transportation network operators often lack an understanding of the value of open data to improve their performance and may not have the expertise or financial resources to implement open data programs. They also face challenges in managing public reactions and expectations about changes in the transportation system that arise from use of open data—what is the best way to tell riders that their customary bus stop is moving?

Privacy and confidentiality concerns span all uses of open data, but in transportation, the ability to track someone's location introduces additional risk. Learning where executives and bankers are traveling or where freight is headed can expose secret negotiations or new strategic initiatives. Broadcasting data about how many riders are commuting from a certain bus stop could give criminals leads on where homes are empty and vulnerable during the day. Data about personal travel could reveal details that a consumer would prefer to keep private and that, if disclosed, could cause embarrassment or financial harm.

IMPLICATIONS FOR STAKEHOLDERS

Based on the opportunities and the obstacles we have outlined above, stakeholders in transportation could consider the following approaches and actions.

- **Transportation providers.** Airlines, railroads, and marine shippers can continue to invest in collecting data and measuring performance, both to improve operations and as a means of differentiation in the market. These data can help identify and remedy variations in operations that harm performance. As carriers open up their data, they will need to establish clear protections for the privacy and confidentiality of their customers and passengers—and communicate these rules clearly.

  Carriers may also seek to inform policies about how regulators gather and share information about carrier performance. By persuading regulators to refine the metrics firms are required to collect, they can eliminate regulations that inadvertently incent operators to behave in a way that negatively affects the customer experience (for example, boarding passengers on planes to meet "on-time" departure criteria even when there are known takeoff delays). Carriers may need to develop new measures and characterizations of data. For example, customer segments are regularly tracked, while complete end-to-end trip data are less common. Airlines and other carriers face a challenge in providing travelers and shippers with easy-to-understand, meaningful metrics on which to base transportation decisions. In addition,

many companies still lack both the skills to make the most of open data and a culture that is prepared for a more data-driven future.

- **Airport and seaport operators.** Opening up performance data can help ports win more customers or at least build better relations where passengers and shippers have no choice. This could include posting average travel delay time today compared with similar days in the past based on weather, events, and other variables. Sharing information with local businesses, such as the number of passengers arriving each year from a certain country, can help the local hospitality industry. Airports can play a role in filling gaps in the data that are available to travelers, such as gate-to-gate times, rather than flight duration times. Shipping port operators can use open data on expected arrivals, backlogs at neighboring ports, and weather conditions to help optimize their operations and ensure they offer valuable service to marine shipping clients. Port operators might also have to enhance their technology and talent base in order to better manage and release their data.

- **Customers.** Passengers and freight customers stand to reap the largest benefits from the use of open data in transportation. Customers can become more aware of data that can lead them to better choices, such as leaving the car at home when it is clear that another mode of transportation will be better or traveling at off-peak times to benefit from lower toll pricing. To make the most of open data, users may have to invest time and effort in mastering the different sources of data and the tools to use them. They may also need to be prepared for the volume of information that may come through the open data "fire hose." Passengers and freight customers have legitimate concerns about how information about their movements and the movements of their goods can be used when it becomes more open. Customers can help shape how the data are used by participating in dialogues with regulators, elected officials, and the travel and transportation industries.

- **Government.** Given the number of different stakeholders involved, the economic impact, and the number of people and businesses the transportation industry touches, regulators and policy makers have much to consider as open data enters the picture. They will need to establish policies about what metrics to collect and share, as well as what incentives to offer to promote desired outcomes for all stakeholders. Regulatory agencies also have a role to play as open data providers and should have clear objectives for sharing data and an equally clear understanding of any risks. As users of open data, regulators and government agencies may need to enact governance rules to allow evidence-based decision making (e.g., using cost-benefit analyses) rather than intuition-based processes, but which also provide protection against politically difficult changes, such as reducing routes to underutilized locations. Policy makers can also consider novel regulations that use open data to increase operations efficiency when infrastructure is constrained. This could include pricing takeoff slots at crowded airports through auctions or enacting dynamic congestion pricing on roadways. Policy makers can advance the use of open data by ending confusion over which public entity has jurisdiction over what areas and making clear what laws apply in specific regions. Such transparency could highlight unnecessary complications and identify ways to make global rules and decisions about data use, rather than adopting less coordinated regional and local ones.

- **Third-party developers.** The energy and talent of third-party developers is already helping to bring out the value of open data in transportation, with bus-tracking tools and other transit apps. Third parties can play an expanding role, by investing in research to understand the data most likely to have an impact on consumer behavior and help to design easy-to-read reports and communication tools. With their focus and expertise, independent developers can also fill gaps in transportation industry capabilities. The core competencies of a transit agency may not include designing uses for open data. With the right data in hand, third parties can do the analysis to determine where new bus routes or train stops are needed, so agencies can focus on making the buses and trains run on time.

3. Open data in consumer products

Every year, $14 trillion in consumer products are designed, manufactured, and sold worldwide, through a set of transactions that involve manufacturers, retailers, and consumers. Walmart alone deals with 60,000 different manufacturers and tracks more than four billion transactions annually. The consumer products ecosystem also includes suppliers of raw materials, government regulators, and third-party data providers. The range of products includes fast-moving consumer goods such as personal health items and packaged food, semi-durable goods such as apparel and accessories, and durables such as appliances. In this chapter we focus on physical consumer products, not consumer services, although we recognize that many of the applications of open data that work in consumer products can also be applied in service industries such as restaurants.

In recent years, many retailers and manufacturers have been facing margin pressure as a result of increased price transparency (thanks to the Internet), and the efficient supply chains of large, low-margin retailers such as Walmart and Amazon. As a result, retailers and manufacturers have been among the most enthusiastic adopters of information technology (see Box 6, "A brief history of data innovation in retail"). Lately they have embraced big data analytics, a set of techniques that MGI estimates could significantly increase the competitiveness of individual companies through more precise insights about consumer sentiment, finer segmentation, and a greater ability to manage operations in real time. [36]

Box 6. A brief history of data innovation in retail

With the introduction of the Universal Product Code system in 1974 and the adoption of computerized cash registers, the consumer products industry began its long history of gathering and crunching data. Initially the data were used for supply-chain management and inventory tracking. Loyalty card programs first appeared in the United Kingdom in the early 1980s and provided retailers with the ability to connect point of sale data with individual shoppers. This gave retailers a way to begin to predict buying behavior.

In the late 1980s, Walmart introduced two critical data-management innovations: an electronic data interchange format that enabled direct links between the computers of suppliers and retailers, and "Retail Link," a system that gave suppliers a direct view of Walmart's store-level inventory. In the 21st century, the quest for consumer data has moved to the Internet. Online marketplaces such as Amazon and Alibaba have based their success on continuous collection and analysis of customer data, down to the last click. Brick-and-mortar retailers have been racing to catch up, with some notable successes. Most recently, information from social websites has become the next frontier of customer data, as demonstrated by companies such as Groupon, LivingSocial, Yelp, Foursquare, and Pinterest.

36 See *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, May 2011.

Open data not only can accelerate the impact of big data and advanced analytics, but it also can uncover new sources of value for manufacturers, retailers, and consumers in consumer products. In this chapter we examine how five levers can deliver value across the consumer products domain through the use of open data:

- **Improved product design and manufacturing.** Product designs can be improved by using open data from social media or crowdsourcing to identify product features that are most valuable to consumers and eliminating features for which customers are not willing to pay extra. Sharing open data benchmarks on manufacturing practices can help to identify opportunities to increase the efficiency in production.

- **Efficient store operations.** Retailers can improve store operations by combining open data such as weather and traffic data with historical sales information to fine-tune demand forecasts and reduce inventory costs as well as sales lost to stock-outs.

- **More targeted marketing and sales.** Retailers and manufacturers can better target customers for marketing and sales by using open data from social media or neighborhood demographics. They can customize advertising as well as store layouts, shelving, and displays according to local preferences, leading to top-line growth.

- **Better-informed consumption.** Consumers can obtain substantial benefits from open data through more informed purchasing. In addition to providing price transparency open data can give customers access to product information—recalls, provenance, social and environmental impact—to help select products that they value more.

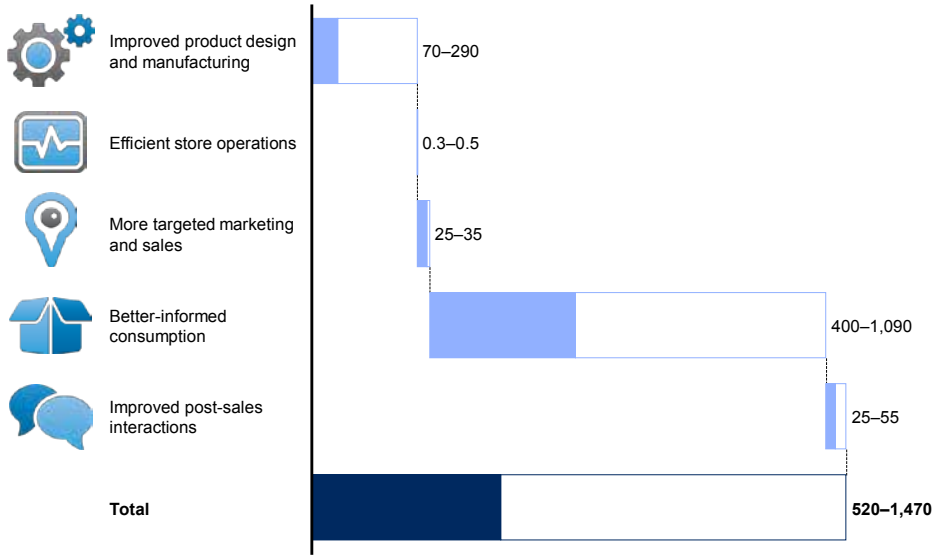
- **Improved post-sales interactions.** Post-sale interactions with customers can be enriched through sharing data on post-sale complaints and resolutions, such as reports about product defects from regulatory agencies or customer service inquiries at call centers. This can enable customers to resolve issues faster, reduce the cost of customer service, and create cross-selling or upselling opportunities.

Together, we estimate that these applications of open data in consumer products can enable $520 billion to $1.47 trillion in value annually.[37] Much of this value will lead to greater consumer surplus from improved transparency into price and product information (Exhibit 5). Market share shifts could also occur across the industry, as companies gain competitive advantage by incorporating open data into their analytics.

Exhibit 5

Five levers can help unlock $520 billion to nearly $1.5 trillion in value per year in consumer products

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

Realizing this value will require overcoming obstacles such as lack of capabilities, privacy concerns, and weak understanding of the business cases for sharing data among manufacturers and retailers. In some cases, such as in open data used for price transparency, data sharing might not be in the interest of manufacturers and retailers, even though it provides benefits to consumers and potentially societal benefits, too.

Many potential sources of open data can be applied to the manufacture, sale, and use of consumer products. They range from information about weather and crops to inform the actions of packaged food suppliers to data about pricing that help consumers make purchasing decisions (see Box 7, “Sources of open data in consumer products”).

37 We capture growth opportunities for the market in the short term or long term, as well as efficiencies gained. Value created is reflected in these estimates, although we also highlight and present estimates for market share that could be captured by leading firms.

Box 7. Sources of open data in consumer products

- **Retailers.** Sophisticated retailers have records of every transaction from point of sale systems (cash registers) and online checkouts. POS data provide insights into the relative performance of categories, products, and stores; when viewed at the shopper level, such data can be used to tailor future offers. Retailers are increasingly able to gather data about behavior in stores, such as the paths shoppers take, and through online clickstreams—a record of all the pages and items a shopper clicked on.

- **Manufacturers.** Manufacturers collect data about consumer behavior and sentiment via traditional methods such as surveys and focus groups, as well as through social media (e.g., Facebook "likes," comments on Twitter). Manufacturers also have large troves of data generated in the management of their increasingly complex supply chains.

- **Government.** Nations around the world furnish demographic and population data, statistics on crop yields, economic reports, patent filings, traffic data, and weather data. Governments also have information on the flow of international goods through ports. Regulatory agencies hold information about product safety and quality.

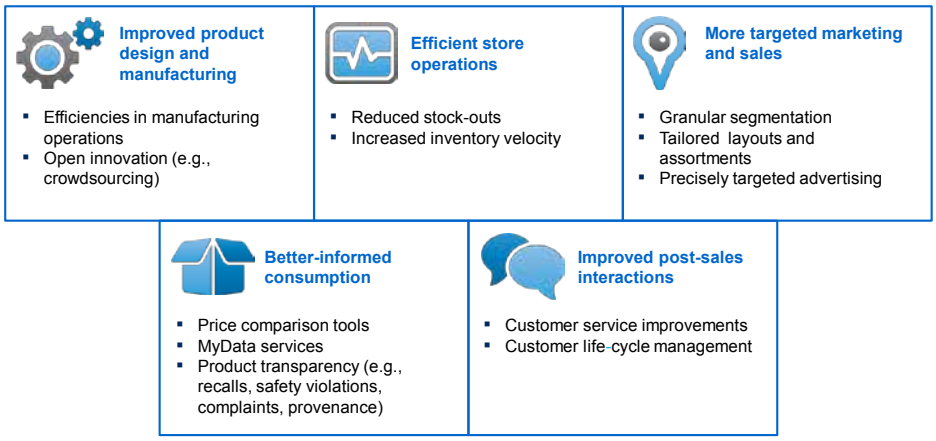
- **Third-party data brokers.** Third parties aggregate all sorts of useful market-wide information, such as market shares, forecasts, and manufacturing efficiency benchmarking metrics. Companies such as Nielsen and IRI aggregate point of sale information from multiple retailers and sell it to retailers and manufacturers. iSuppli conducts teardowns of electronics products, providing the entire industry with a view of the components used and their costs. Payment processors such as credit card companies collect transaction data. Companies such as Quality Assurance International certify that products that claim to be organic actually are.

- **Consumers.** In addition to leaving their tracks in online and offline stores, consumers volunteer valuable information in surveys and focus groups, by engaging with brands on social media, and by joining crowdsourcing efforts such as polls and contests. Some retailers and other organizations that interact with consumers provide MyData to consumers—data about an individual's shopping habits, transactions, and other activity.

FIVE LEVERS FOR CREATING VALUE IN CONSUMER PRODUCTS

We identify five levers for creating value in consumer products. These levers can create value by enabling performance improvements and greater efficiency across the value chain, from product specifications to marketing and sales, and post-sale support. In addition, open data can be harnessed by consumers to make better purchasing decisions. For each lever, we identify major strategies and approaches (Exhibit 6).

Exhibit 6
How five open data levers create value in consumer products



SOURCE: McKinsey Global Institute analysis

Improved product design and manufacturing

Open data can help manufacturers design and build products that are better matched to customer requirements, are of higher quality, and are produced more efficiently. Using openly accessible social media data, manufacturers can develop insights about the relative importance of different product specifications. They can also learn about safety issues when data on safety recalls or complaints submitted to government regulatory agencies are made open.[38] Manufacturers can get input on design requirements directly from their customers by opening up platforms for crowdsourcing. Additionally, manufacturers can use shared data to stimulate efficiency gains in manufacturing processes.

Crowdsourcing—opening up the design process to input from the public—can improve quality of products by incorporating more ideas, creating designs that are more responsive to customer needs, and potentially making R&D more efficient.[39] Danish toymaker Lego, for example, introduced a software application called Digital Designer that allows users to publish Lego designs online and purchase the parts to build them. Consumers can also see designs that other Lego fans submitted, leading to a cycle of continuous improvement on previous

38 For example, in the United States, the Consumer Product Safety Commission publishes multiple data sets about product recalls. The agency also maintains the National Electronic Injury Surveillance System, which samples emergency room data to estimate the number of injuries associated with consumer products.

39 Crowdsourcing can incorporate open data in two ways. Open design allows designers to see each other's work and respond; participants are motivated by the potential approval of others. Another approach is posting data and inviting participants to compete by extracting insights. Netflix and Goldcorp have opened significant amounts of data to elicit crowdsourced insights.

designs. By using open design platforms, manufacturers can gather more useful inputs to inform design of products and increase the effectiveness of in-house design efforts.

Additional value creation opportunities come from sharing data on manufacturing performance. By using shared operational data, such as utilization, yield, and scrap metrics, manufacturers can identify areas for improvement and initiate programs to eliminate waste and reduce costs. Some companies are already using big data captured through networks of sensors in factory machines (an example of the growing Internet of Things) to raise efficiency.[40] Such benefits can be multiplied when these operational data sets are shared by multiple players. Today, when these kinds of data are shared, it requires a good deal of human labor to aggregate reporting from different companies. With sensor-enabled "digital factories," much of the data collection and aggregation could be automatic and could be performed at close to real-time speeds.

Several efforts are under way to develop data standards that are crucial for capturing efficiency gains from open and shared data. For example, MTConnect, a set of royalty-free manufacturing standards, was created in 2008 in an attempt to standardize the reporting of data from digital manufacturing processes. The Smart Manufacturing Leadership Coalition, launched in 2012, aims to encourage data sharing by standardizing and aggregating industrial data to identify better manufacturing processes. One specific goal of the group is to standardize data reporting across multiple types of sensors and to unify data reporting across open platforms.

Another type of data that can be shared is teardown information. Independent companies dissect products such as consumer electronics to identify the component parts and costs. These third parties act as data brokers for aggregated data on product teardowns, which can yield insights into design features that reduce manufacturing costs or inform pricing.

40 For more about big data and advanced analytics applications in retail, see McKinsey Global Institute's *Big data: The next frontier for innovation, competition, and productivity*, May 2011, and *Disruptive technologies: Advances that will transform life, business, and the global economy*, May 2013.

We estimate that open data can help to enable $70 billion to $290 billion in annual value through improved product design and manufacturing. This value is mostly derived from the operational improvements and cost savings that manufacturers generate through use of operational benchmarks to find cost savings.[41] Additional value can be captured through crowdsourcing to reduce development cost and times.[42]

Efficient store operations

Using such open data as weather and demographic information supplied by government agencies can help retailers fine-tune store operations. The UK grocery chain Tesco, for example, has been combining weather data from various sources, including the UK government and a third-party data provider, with sales records across thousands of stores to create hour-by-hour demand models. Given the strong influence of weather on consumer activity and the great variability in weather in the United Kingdom, Tesco found that it was important to model weather in as much detail and as frequently as possible.

Using detailed weather data in its demand models allows Tesco to adjust the product mix at the individual store level, based on weather data and past purchasing patterns. Tesco calculated that an 18-degree Fahrenheit rise in temperature typically corresponds to a 300 percent increase in sales of barbeque meat and a 50 percent increase in sales of lettuce, suggesting that stocking more meat and lettuce when higher temperatures are forecast is a smart investment.[43] Such modeling has helped Tesco reduce revenue losses caused by stock-outs and the amount of inventory that Tesco holds, reducing losses to spoilage.

We estimate that if open data were widely adopted across the fresh foods category in grocery industry, it could enable potential value creation of as much as $300 million to $500 million annually in efficient store operations. This value would be seen in tighter inventory management, resulting in reduced waste from spoilage. Additionally, we estimate that reducing stock-outs could contribute $70 billion to $110 billion annually in revenue, but most of that value will be realized as share shifts among retailers, which we do not count in our estimates. There are also other potential applications of open data that could improve store efficiency. Stores can manage labor more efficiently by using data on traffic, weather, and events, and they can adjust store inventories, based on data about product placements on television shows. Retailers can also optimize real estate costs and location selection by using local traffic data, inventory of nearby shops, and patterns in utility costs.

41 Typical manufacturing cost improvements from implementing benchmarked lean manufacturing processes range from 15 to 25 percent over the course of 18 to 24 months (*McKinsey analysis).

42 We did not estimate these benefits directly, because of the highly variable nature of R&D processes in different firms and industries.

43 Julia Werdigier, “Tesco, British grocer, uses weather to predict sales,” *New York Times*, September 1, 2009.

More targeted marketing and sales

Open data approaches can help manufacturers and retailers target individual consumers and customize offerings for specific segments. Granular demographic data allow retailers and manufacturers to optimize merchandising on a store-by-store basis, customizing shelving, pricing, assortment, and bundling for each outlet. Also, by tapping into increasingly granular data sources, such as traffic flows and social media buzz, retailers and manufacturers can deliver more relevant advertising to individual consumers.

Historically, the most important source of information in the consumer retail arena has been purchase data collected at the point of sale. These records reflect the actual purchases that are made and are the most accurate view of real-time demand. This information is used for demand planning, pricing, and a variety of other uses. As a result it has been considered a source of proprietary advantage and not shared.

However, by sharing point-of-sale data under limited circumstances—avoiding sharing with direct competitors—companies can generate insights and marketing approaches that they could not get to with proprietary sales data alone. Retailers have shared data in a limited way, by allowing commercial data brokers such as Nielsen and IRI to collect data from across the consumer goods industry and sell aggregated data back to retailers and manufacturers. Sharing is also seen in loyalty programs. Nectar, a UK-based loyalty card, can be used at Sainsbury's for groceries, BP stations for gasoline, and Hertz for car rentals. Sharing aggregated consumer data allows the three companies to gain a broader and more complete perspective of consumer behavior, while preserving competitive advantage by sharing only with retailers in non-competing industries. By forming coalitions, firms are able to gain insight into a broader share of wallet of their consumers. Often, such granular, customer-level data are available only at a high cost or not available to all parties.

Retailers are starting to combine proprietary or high-cost data with additional sources of data to devise new approaches to customers. To target consumers more precisely, retailers are combining sales information they have accumulated (either by themselves or through partnerships) with highly detailed public census and demographic data. Based on localized demographic data, retailers have been able to capture market share by adjusting store formats, product assortments, and displays to specific neighborhoods. This has the additional benefit of delivering products better fit for consumers at each store location.

The introduction of additional data is also helping to finely tune advertising. By taking customer information that they already have and combining those data with novel sources of open data such as social media chatter, local weather, and any MyData that consumers share, retailers and manufacturers can deliver more relevant and timely advertising and get more out of their advertising budgets.

Through more precisely targeted marketing and sales, we conservatively estimate that open data could help create $25 billion to $35 billion in value annually.

Better-informed consumption

Open data can help consumers make better purchasing decisions by enabling greater price transparency and improving access to additional product "metadata" that allow consumers to make purchase decisions based on other factors they value. Additionally, MyData can offer value to consumers through increased convenience, information about their behavior, and an opportunity to monetize their own data.

Online shopping has already brought price transparency to virtually every consumer product category. Lately, brick-and-mortar stores have been forced to respond to additional sources of transparency as smartphone apps have made it possible for consumers to check prices across the Web (and at other physical stores) while they are shopping in a store. RedLaser is a mobile phone application that allows a customer to scan a product barcode in the store and compare prices at other stores and on online retail sites, while also getting additional product information, such as potential allergens in packaged food.

While online price comparison tools have already had significant impact, we believe that much of the value from price transparency has yet to be captured. Only about 10 percent of consumers currently use price-check applications on their phones, and price-comparison services are just now reaching major categories such as groceries. MySupermarket, a UK-based website for online grocery shopping, says it saves customers $26 per shopping "trip."

Price is not the only shopping information that consumers seek. They are often willing to pay premium prices for products that they perceive to be of higher quality or that are proven to be safe. Since melamine in Chinese baby formula caused six deaths and sickened more than 300,000 babies in 2008, Chinese parents have paid more for certified safe milk products or imported their own from overseas. The rise of stores featuring organic and all-natural products is another indication that consumers are increasingly willing to pay premium prices for products with additional certifications.[44]

Open data is now playing a role in such buying decisions, thanks to a growing number of non-profits and third-party companies that are providing information about how products are made and the purity and provenance of their contents. Organizations such as Quality Assurance International certify that products are organic or grown under sustainable conditions, while websites such as Sourcemap provide crowdsourced data (from both manufacturers and consumers) about the supply chains and environmental footprints of various products. This additional product information, if provided in an openly accessible, machine-readable format, can allow consumers to shop for products that meet their social and ethical standards and get increased satisfaction from the goods that they purchase. It also allows manufacturers and retailers that can meet those standards to charge a premium for products.

MyData can help customers in several ways. First, data on previous purchases can offer convenience, for example by guiding consumers to supplies or replacement parts, based on purchase history. Second, customers can also use MyData to get insights into their own behavior, learning for example what percent of their grocery purchases are high in fat and receiving offers of personalized

44 Kristen Park and Miguel I. Gomez, "Do price premiums exist for local products?" *Journal of Food Distribution Research*, volume 43, number 1, March 2012.

diet plans based on that data. MyData can also empower consumers to correct, improve, and otherwise control the data collected about them.

Data aggregators are also beginning to offer MyData services. Acxiom, a data broker that collects data about consumers from many open and proprietary sources, has created a portal where consumers can view some of this information. Consumers can correct information such as birth date or household income, suppress specific data fields, or even opt out of Acxiom completely. By sharing data with consumers and relinquishing some control of the data, the firm hopes to reduce backlash over privacy concerns and improve data accuracy, which can increase the value of the data.

Consumers are now getting the chance to sell their own data, too. Personal, a startup based in Washington, DC, allows consumers to sell their own information in an online marketplace where consumers and marketers meet, enabling companies to purchase detailed personal data such as food preferences or financial information provided by consumers.

Through more informed purchasing, we estimate that open data could enable $400 billion to nearly $1.1 trillion annually in value for consumers. This value was estimated based on the price premiums consumers are willing to pay for products that have certain attributes that open data confirms. This premium ranges from 10 to 80 percent.[45] In the long run, this lever also provides net societal benefit, as shoppers are able to purchase differentiated products that suit their needs at the lowest possible price, while producers are able to compete more effectively through differentiated products that better match customer needs. Additional value can be created through greater use of MyData, but we do not attempt to estimate how much, given the early stage of development in MyData use.

Better post-sales interactions

Post-sales interactions were once characterized by returning defective products, mailing product registration cards, and phoning customer service centers. Today, post-sales interactions are becoming part of customer life-cycle management, an important input into product design, and a source of additional sales. Releasing data from customer service interactions on other customers' issues and resolutions can help customers rapidly solve their own problems, thereby increasing satisfaction and reducing the cost of customer service. Similarly, companies can use open data about a customer (e.g., demographics, social media behavior) to help tailor the interactions, while also providing additional valuable offers. Finally, collecting publicly available information about complaints can help companies and governments recall defective products more quickly.

There are many potential benefits from making data on customer service interactions accessible. Customer service already has become more communal and interactive, through the use of publicly accessible and searchable online forums and social media channels. BestBuy, for instance, says it reduced customer service complaint volume by 20 percent after sharing the responses to common complaints on Twitter and crowdsourcing solutions from its employees. Intuit TurboTax has started to display related user-generated questions and answers adjacent to its online help pages, improving the user experience of its site and reducing customer service costs. Mining the data available about

45 We estimated price premiums by examining price premiums for a variety of products and certifications, such as Fair Trade coffee and organic produce.

customer service interactions can also yield useful insights on various topics from consumer preferences to pain points or product quality.

Third parties are also getting involved in post-sale interactions by using open data. One open data application involves replacement parts, an important aspect of after-sales interactions. Thingiverse creates replacement parts for a variety of products using 3D printing. Instead of trying to hunt down parts from the manufacturer or a repair service, customers can create their own replacement parts based on specifications that have been crowdsourced to the open Thingiverse platform or uploaded by the manufacturer.

Open data also can help reduce the cost associated with defective products, which in the United States alone are associated with estimated economic losses of $900 billion a year through injury, death, and property damage. Governments, manufacturers, and retailers may be able to detect product issues more quickly by harnessing multiple open data sources such as social media interactions, customer service records, complaint data submitted to consumer groups and regulators, injury data from hospitals, and search-engine traffic. Web search data have already been shown to be helpful in quickly identifying the vector of flu outbreaks and could be used to detect product safety issues, too. Once a recall is issued, open data can help spread the word. For example, by linking recall data to UPC codes, consumers could be alerted at checkout about recalls or safety messages related to their purchases.

Through improved post-sales interactions, we estimate that open data could help create value of as much as $25 billion to $55 billion annually, primarily from reduced costs of customer service.[46] We do not attempt to quantify revenue opportunities from targeted cross-selling and upselling, which is described in the section above on “More targeted marketing and sales.”

BARRIERS AND RISKS

Capturing the potential value that use of open data can help enable in consumer products will require overcoming some significant obstacles and addressing perceived and actual risks. These issues range from privacy and data security concerns to lack of understanding about potential benefits, to technical and capability gaps.

Consumers, manufacturers, and retailers all have concerns about the security and privacy of information that would be collected and used in open data applications. Shoppers fear that the information that online and offline retailers collect about them from shopping interactions and their behavior on social media could be used to violate their privacy and even result in harm, such as a lower credit rating. Manufacturers and retailers have reservations about sharing proprietary data in an open fashion to support industry-wide benchmarking and other performance-improvement efforts. They are concerned about the data somehow being associated with their organizations and about how data will be shared and re-used.

Another barrier is a lack of understanding among retailers and manufacturers about the value of sharing data with other companies. Companies are concerned

46 We applied a 20 percent reduction in complaint volume to all retail, taking into account the varying rates of complaints by key product types (e.g., high customer service costs for personal technology products and low costs for groceries).

about helping competitors learn about their sources of advantage (or exposing weaknesses). Similarly, retailers and manufacturers are concerned that shared data could benefit one party over the other in terms of who captures margin. Finally, companies are not yet convinced of the value to be gained in sharing data with consumers.

We also see technical barriers and gaps in capabilities that make it more difficult for a consumer products company to embrace open data. There is a lack of standards for incorporating metadata, such as supply-chain information, into open data sets that are easy to compare across manufacturers. For example, recall data made publicly available by the US Consumer Product Safety Commission starting in 2010 is not currently reliably associated with UPC data, preventing third-party companies from incorporating these data effectively into the information provided to consumers at checkout. Other gaps exist in validating the accuracy of reported product metadata. The sheer volume of information collected makes it quite challenging to uniquely identify customers.

IMPLICATIONS FOR STAKEHOLDERS

- **Consumers** have the most to gain from open data in the consumer products domain. They have the opportunity to take advantage of price comparison tools and other applications that provide information that allow them to make purchase decisions based on what they value about different consumer products.

  Consumers can become more educated about how data are used and what the potential risks are. They can start by reading data policies and making thoughtful decisions about what data they would like to share and considering the benefits and risks of sharing such data: are they willing to have their shopping patterns made available in return for access to lower prices or tailored offers? Are they willing to have those data shared with other parties? Consumers have the opportunity to capitalize on opportunities to access information about themselves through MyData portals, or even request the ability to do so. Consumers could also participate in the public and policy-making dialogues regarding privacy and the use of personal data.

- **Companies** have significant opportunities to combine open data with data from other sources to capture value and enhance their competitive positions. To do so, they will have to find sources of open data and develop the technical and organizational capabilities to analyze and make better decisions based on those analyses. In some cases, companies will have to determine if there is a way to share data with competitors or with partners and customers across the value chain (as in retailers collaborating with manufacturers) in ways that benefit all parties. To take full advantage of open data, companies will have to agree upon standards that make data from multiple sources comparable and easy to integrate.

  In addition, companies will need to make a concerted effort to communicate the value proposition of open data use to consumers. Companies can be transparent about their practices for data collection, storage, use, re-use, and sharing by publishing these practices in plain language. They can also choose to make it easy for consumers to access and potentially correct data about themselves and express preferences about the use of data about themselves through MyData systems.

When sharing data, companies need to establish best practices to ensure data security and the appropriate amount of anonymity, for either personal or corporate data. Companies can designate trusted third parties and set guidelines and standards for data use and re-use as well as for sharing.

- **Government** plays an essential role in weighing the benefits of open data and the need to protect privacy and property rights. Governments may need to consider regulations to control the collection, storage, use, and/or sharing of data and to establish avenues of recourse if data are misused. These rules could cover secondary use of data (for example, what would be an appropriate use of consumer purchasing data for underwriting?).

  In some cases, governments could encourage or mandate the release of open data when it could serve broader, societal goals or help inform consumer decision making, such as releasing information to promote transparency about price, provenance, and potential health impacts. Government can also lead efforts to set standards for data measurement and formats to make data easier to use and to accommodate new types of data, such as environmental data.

4. Open data in electricity

Global demand for electricity is projected to be twice as high in 2030 as it was in 2000. There are many challenges that accompany this growth. Generation efficiency has not improved significantly in decades, nor has grid efficiency; in 2011, an average of 8.1 percent of generated electricity was lost in transmission and distribution around the world, including through pilferage.[47] New sources of energy coming onto the energy grid, such as distributed generation from rooftop solar panels, are increasing operational complexity. There is also rising pressure from the public and government to limit the effects on pollution and climate change from the generation and use of electricity, raising interest in technologies and actions to conserve energy. This would entail cutting residential consumption, improving energy efficiency of homes and buildings, and getting utilities to invest in smart grids, adopt renewables, and accept new regulatory frameworks. In developing countries, demand growth significantly exceeds supply, because of insufficient capital investment and high costs of connecting remote areas to the power grid.

Open data can help address many of the issues facing the electricity industry, its customers, and government regulators. Making data more open through industry benchmarks can help utilities save on capital investments and operations. As utilities experiment with smart grids and other new technologies, sharing data about technology pilots can avoid duplicated efforts and optimize investment. Sharing data such as technical standards, work processes, task time benchmarks, construction plans, performance management systems, and asset replacement schedules can be used to identify areas for improvement, which can lead to significant savings. Additionally, providing data to electricity users on their usage patterns and showing comparisons with usage by similar customers, can lead to more efficient electricity consumption. While these types of data can help to drive savings, significant investment in technology and operations will be required to fully capture this value.

In the long run, data sharing could help the economic and environmental viability of the energy industry. Lower costs of operations and investment could lead to a more sustainable economic structure, with a larger share of funding derived from current operations and less reliance on future growth to pay for investments. Transparency about the environmental impact per unit of electricity generated can also begin to shift generation investments, beginning with identifying areas that have the largest potential reduction in pollutants and leading to lower overall environmental impact in the future.

47 Measured by the amount of electricity generated as a percentage of energy fed into the generation process; *Electric power and distribution losses*, World Bank, www.data.worldbank.org.

In this chapter we identify five levers for using open data to improve efficiency across the electricity value chain: improved generation investment, efficient generation operations, optimized investment in transmission and distribution, efficient transmission and distribution operations, optimized retail and consumption.[48]

- **Optimized generation investment.** By using publicly available data such as siting applications, resource planning documents, and regulatory filings, utilities can streamline the very challenging and complex permitting process for new power projects. Capital productivity data from across the industry can help inform investment decisions.

- **Efficient generation operations.** By sharing benchmarking data about processes such as maintenance routines and equipment utilization, power companies can optimize the performance of generating operations.

- **Optimized investment in transmission and distribution.** As with generation investments, open data can help optimize investments in transmission and distribution (the lines and equipment that convey electricity to users and between utilities).

- **Efficient transmission and distribution operations.** Shared open data can help identify efficiency opportunities in transmission and distribution operations. Sharing data gathered via "smart grid" technologies can be particularly helpful.

- **Optimized retail and consumption.** Residential and business users can make better decisions about which appliances and equipment to buy or what electric service to use (where they have a choice) by using open data that makes it possible to compare products and services. For residential and business customers, MyData combined with community-wide data can show customers how their energy use compares with neighbors and peer organizations

We estimate that open data has the potential to add $340 billion to $580 billion of value annually across the electricity value chain (Exhibit 7). By providing visibility into current inefficiencies and areas of opportunity, open data can help set businesses and individuals down the path of improvement, though in some cases, achieving the improvement will require substantial additional resources. The biggest opportunities lie at the consumer end of the value chain, where energy efficiency, price transparency, and better retail options could help enable value of $180 billion to $310 billion annually.

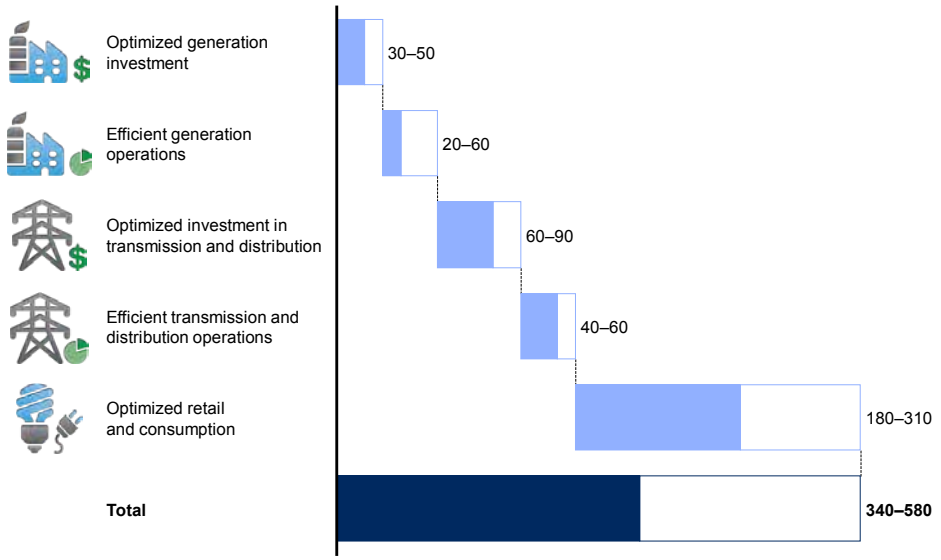
48 We include all entities involved in this value chain (integrated utilities, generators, transmission system operators, distribution system operators, traders, retailers, and end users, among others). We also include generation from all fuel sources, including fossil fuels (coal, oil, and gas), renewables (such as solar, wind, hydro, and biomass), and nuclear power.

There are risks and barriers that stand in the way of realizing this potential value. There is a lack of standards in data collection and, in some places, misaligned incentives. European customers, for example, could benefit from reduced electricity bills as a result of energy-efficiency programs, but their electric companies, whose revenue is tied to consumption and based on regulated electricity prices, have little incentive to invest in programs and technologies that reduce consumption. Privacy is also a significant concern for some consumers, and electricity producers are wary of sharing information that could jeopardize a competitive advantage or reveal unflattering information, such as a poor environmental record. To overcome these challenges, governments, utilities, and consumers will need to agree on a framework for protecting the confidentiality of data, while also creating incentives to encourage data sharing, possibly through regulatory changes and legislation.

Exhibit 7

Five levers can help unlock $340 billion to $580 billion of value per year for electricity producers and users

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

The producers of electricity, their customers, regulators, and third parties produce a wealth of data. Some of this information is open today, and much more can be opened through data sharing (see Box 8, “Sources of open data in electric power”).

Box 8. Sources of open data in electric power

- **Power suppliers.** Generating companies, grid operators, and retailers possess company-level data regarding their business practices and operations. These data sets include performance data as well as information required by regulators such as health and environmental reports, and data for permitting and licensing applications. They also generate results from government-funded pilots for new technologies such as renewables, which are usually publicly accessible. Many utilities are also installing smart meters, which report near real-time data on energy consumption, outages, and power quality. Suppliers regard much of the data they collect as proprietary, including detailed information about utilization and cost metrics, workflows, and management practices that they are not required to submit to regulators.

- **Regulators.** Government regulators, such as the Federal Energy Regulatory Commission (FERC), collect health and safety accident reports, environmental impact assessments, and prudency determinations (the amount of justified capital investment costs that utilities can pass on to consumers through higher rates). They also compile information about rates, returns, and other investment and performance data; these vary among regulatory schemes.

- **Government agencies.** Some government agencies that do not regulate the electrical industry collect macroeconomic data about the energy sector. For example, the US Department of Energy aggregates national statistics about supply, demand, pricing, imports, and exports. It also issues projections and forecasts. Data published by other government agencies, such as meteorological and demographic trend information, can be used for demand forecasting, site selection, and other purposes in the electrical industry.

- **Energy users.** Through their meters and bills, energy users generate a wealth of data. With the introduction of advanced metering systems, data are now collected as frequently as every 15 minutes. Additional data about energy users include demographic data and information about their homes or commercial buildings (e.g., type of roofing, heating methods).

- **Third-party data brokers.** Independent companies in the power sector aggregate data from multiple sources (e.g., generating companies). These data could range from statistics about the efficiency of specific generating plants to the typical costs paid for labor in a particular area. Many third parties act as trusted brokers, with which companies share proprietary operating data that will go into industry-wide databases for benchmarking and other uses. Other third-party firms package performance and cost data for users.

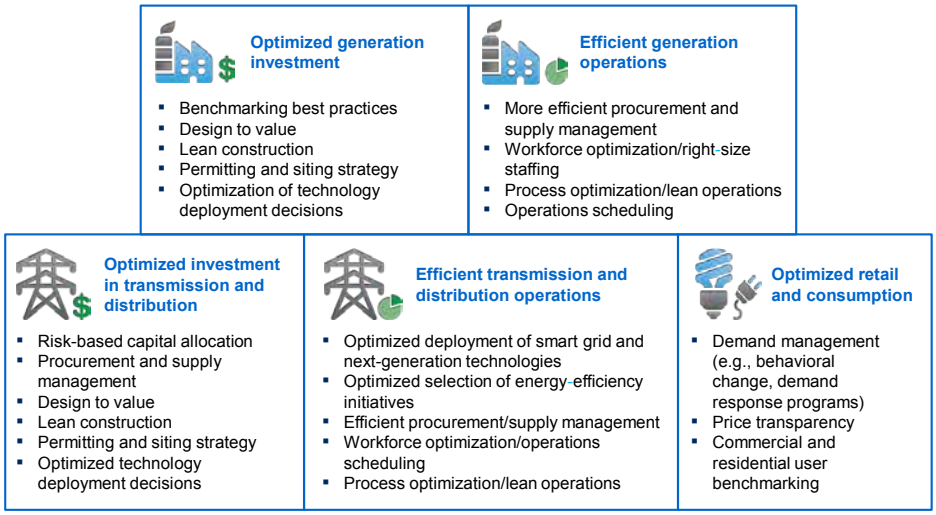
- **Industry groups.** Membership groups such as the World Association of Nuclear Operators collect information from their members for benchmarking and other purposes.

HOW FIVE LEVERS CREATE VALUE IN ELECTRIC POWER

Over the next two decades and beyond, skillful use of the five levers we identify for using open data can bring about significant improvements in how the global electrical industry invests in new capacity and runs daily operations and can influence how its customers use power. Together, these improvements can have very significant economic and environmental impact and enable the industry to meet the growing demands of a power-hungry planet (Exhibit 8).

Exhibit 8

How five open data levers create value in electricity



SOURCE: McKinsey Global Institute analysis

Optimized generation investment

Approximately $5 billion is projected to be spent on new generating capacity globally from 2015 to 2035.[49] Using open data to guide site selection, permitting, technology deployment, and engineering and construction can save time and expense.

Most generating companies already use publicly available data to help inform site selection strategy. For example, historic weather data are crucial in determining where to locate wind turbines and solar panels and in estimating how much power they are likely to generate at any given time. Weather data also help demand forecasting: hot days lead to high air conditioning use, and detailed analyses of weather data help utilities identify areas where demand is likely to exceed supply, which is an important factor in site strategy.

Open data also can help inform decisions about what type of generation capacity to invest in, by providing a more complete picture of total economic and environmental costs. Key considerations for which open data can provide additional insight include choice of fuel, the cost and environmental impact of producing and transporting the fuel, as well as the impact of using that fuel for generation. Such analysis can create a more complete understanding of the costs and benefits of particular choices about what kind of capacity to build and where.

49 *International energy outlook*, Energy Information Administration, May 2013.

Once a choice of plant type and site is made, the next step is the time-consuming process of permitting, which usually involves negotiations with citizen and environmental groups in addition to meeting the requirements from multiple regulatory bodies, including national and regional governments. While data alone cannot solve all the political, regulatory, and technical challenges that determine how and where electricity is generated, open data can reduce friction in the permitting process by creating transparency and promoting collaboration among stakeholders. By aggregating siting best practices, detailed accounts of previous negotiations, and numerical data regarding the specific details of each individual case, such as proposed capacity, location of proposed site, and type of generation, stakeholders can come together to streamline the siting and permitting negotiations process.

Openly sharing benchmarking data and related best practices for construction can help utilities bring new capacity online more efficiently—saving on materials and components, and avoiding delays and cost overruns. Using open data on procurement and supply management and design to value data (for example, industry design benchmarks and data about peer installations) can give utilities money-saving insights into how to design a project and engage contractors. When a project is under way, data-driven project management techniques based on open data such as benchmarks for tool utilization rates can help control costs.

We estimate that the use of open data can create opportunities to save $30 billion to $50 billion annually in the design, engineering, procurement, and construction of power generation facilities. This represents a 15 percent savings in capital expenditures for new plants in 2013. We do not attempt to quantify the impact of open data on improved siting and permitting processes, which vary widely among regulatory regimes.

Efficient generation operations

Sharing industry data about workforce management, operating processes, and procurement practices can help drive efficiencies across generating operations. The shared data sets that enable operating improvements range from time required to perform specific maintenance tasks, to equipment utilization rates, and typical costs for stocking maintenance parts. Additionally, sharing information on scheduled maintenance and downtime among power generating companies could help management prevent interruptions in supply. We estimate that sharing open data could enable $20 billion to $60 billion in savings annually in generation operations. This represents approximately 20 percent of what was spent on global power generation operations and maintenance in 2012.

Optimized investment in transmission and distribution

The power grid (an electric company's transmission and distribution system) transports power from where it is generated to where electricity is used. Open data can help improve the development of safe, affordable, and reliable grids. Shared data on siting and permitting, equipment mortality curves, and historical and projected capacity, as well as benchmarking, can hasten deployment of grid investments and reduce costs.

We estimate that sharing open data has the potential to uncover value of $60 billion to $90 billion per year through more efficient use of capital in building transmission and distribution (T&D) infrastructure. Use of benchmarks for procurement, construction, and other metrics (similar to those described for generation investments) and the implementation of associated best practices, can lead to 10 to 15 percent savings in T&D investments. The impact of construction benchmarking and best practices would be somewhat smaller in value than in power generation because transmission and distribution infrastructure is less capital-intensive and can be deployed in small increments. Additional value can be captured through risk-based capital investment, where proprietary information on a utility's asset health can be combined with equipment and mortality curves and historical and projected capacity to ensure that each marginal dollar of investment is providing the greatest benefit in terms of the reliability of the system. Improved siting and permitting can unlock value in T&D as well, although we do not attempt to quantify the potential impact due to the variation in regulatory requirements across jurisdictions.

Efficient transmission and distribution operations

Sharing T&D operating data across the industry can help utilities identify improvement opportunities in procurement and supply management, workforce management, and operations scheduling. Transmission and distribution operations also can be made more efficient through grid control technologies. For many years, the power industry has been assessing the value of investments in smart grid technologies that can monitor and control electric use and can automatically route power around failed lines or transformers. But widespread adoption has not occurred for a number of reasons, including a lack of data. Sharing standardized data on the results of smart grid deployments and trials can help operators analyze the impact of different next-generation grid control technologies, which can lead to greater confidence in investment decisions and accelerates adoption. Grid operators can also share data on purchasing and installation, further optimizing and accelerating deployments. Open data also has the potential to accelerate the timeline for realizing savings from energy efficiency, reduced outages, increased transmission throughput, and lower inspection and maintenance costs.

Across transmission and distribution operations, we estimate that open data can help enable $40 billion to $60 billion in value annually, the majority of which comes from operations and maintenance cost efficiencies. We also recognize the significant potential for open data to help accelerate the deployment of smart grid technologies such as grid control, which we have not sized directly.

Optimized retail and consumption

We estimate that helping energy users make smart decisions about how to use electricity will be the greatest source of economic impact from the use of open data in the electricity industry. Through demand-management programs enabled by smart grid technologies and other energy-efficiency programs, open data can be used to help modify the behavior of residential and commercial customers, allowing utilities to postpone capacity additions, reducing costs to users, and limiting the growth of greenhouse gas emissions. Open data has a direct application in demand-side management—inducing customers to use less electricity.

Demand management programs can take many forms, including informational programs to drive behavior change, demand response programs, and incentives for targeted energy efficiency investments, among others. Opower, a US-based service, works with utilities to benchmark individual customer usage (provided as MyData) against that of neighbors. The information is sent with the customer's bill and includes personalized tips on how to reduce consumption. Demand management programs take this one step further, giving consumers real-time information on systemwide demand so that they can reduce use when wholesale prices are high or the system reliability is compromised. These programs can rely on consumers' good will to reduce usage when demand is high or rely on economic incentives, with consumers bearing higher rates for consumption during peak demand. Additional economic incentive can be created to encourage the installation of more energy-efficient products (e.g., tax deductions for installing double-glazed windows).

Open data sets have been harnessed to inform consumption for many years. Since 1992, the US government has provided open benchmarking data for industry through the Energy Star Energy Performance Indicator tool, which provides benchmarking data on energy use within several industries. US iron and steel manufacturers report saving from 9 to 18 percent in electricity costs by using the system.[50] New York City released energy consumption data about large commercial buildings in 2011 and recently began reporting energy use of residential buildings as well. By bringing inefficient buildings up to the current median of their peer groups by catalyzing energy efficiency efforts through open data benchmarking (where the block-by-block consumption data are made publicly available), the city expects to reduce commercial energy consumption by 18 percent and cut emissions, too. Similar open data about building energy use has also been made available in Europe.[51] The Energy Star program also rates and certifies consumer appliances and posts data publicly so shoppers can consider energy efficiency in their purchase decisions.

Additional customer savings can come from using open data sets that allow price comparisons among electric retailers. Germany, the United Kingdom, and the US state of Texas, among other jurisdictions, allow resellers to compete with electric utilities to sell service to users. Check24, a service that operates in Germany, and energyhelpline.com in the United Kingdom provide comparison data on competing electricity retailers. Customers save an average of $170 to $250 per year by comparison shopping across electric retailers, according to the services.

We estimate that through energy efficiency and price comparison opportunities, open data can help create value of $180 billion to $310 billion annually. Customer benefits include approximately $70 billion from reducing building electrical consumption (such as lighting and air conditioning), $2 billion to $20 billion from adopting more efficient electrical products, $30 billion to $80 billion from purchasing decisions based on price comparisons across electrical retailers, and $59 billion from the use of smart grid demand-management technologies.[52]

---

50 See www.energystar.gov.

51 We estimate 3 $GtCO_2$ of abatement potential globally in the building sector, from both electricity and fuel.

52 See *US smart grid value at stake: The $130 billion question*, McKinsey & Company Electric Power and Natural Gas Practice and Business Technology Office, 2009. The $59 billion estimate is based on the potential value of using open data to accelerate adoption of smart grid technologies.

BARRIERS AND RISKS

Many obstacles stand in the way of realizing the full potential of open data to help create value across the electricity ecosystem. These barriers involve technical and regulatory issues as well as gaps in knowledge and capabilities.

Among the greatest obstacles is the reluctance of players across the value chain to share data. Many companies in power generation and in transmission and distribution harbor concerns over confidentiality and the possibility that the proprietary data that they agree to share for benchmarking purposes will be misused. Commercial and industrial energy users have similar concerns, as do residential consumers. Some customers fear that the monitoring that is made possible by advanced metering systems is an invasion of privacy and possibly a governmental intrusion. Consumer privacy concerns thwarted plans by Oxxio, a Dutch energy company, and the Dutch government to roll out smart meters to all seven million households in the Netherlands. In addition to privacy concerns and worries over government intrusion, consumers pointed out that criminals could obtain usage data and determine the optimal time to attempt a burglary.

Current business models and regulatory frameworks are not necessarily aligned to create value from open data in the electricity domain. In many markets, overlaps in regulation between national and regional bodies often lead to confusion and sub-optimal investment in electricity generation. Some government programs to incent energy conservation also have unintended consequences: by defining a discrete target for reduction, utilities may make the investments in technologies and programs only to hit a specific target, rather than implementing a program that would lead to the greatest and most cost-effective conservation. Strict regulation over siting and permitting for new generation as well as the limits on rates that can be charged for electricity can also lead to inefficient investment decisions.

Technical issues could also impede adoption of open data. For example, electricity data are not measured or reported using standard methodologies and formats. Regulators in different areas require different formats, hampering comparability. Governments also vary in how they calculate carbon impact, which affects benchmarking of generating investments, and there are no standards for reporting data about trials of new smart grid or generating technologies, leading to a lack of confidence in the results and potentially delaying investment and innovation. In addition, when data are available, information is not always presented in easy-to-understand, standardized views or summarized in reports that can help people grasp the meaning of the data and analyses.

Finally, there are gaps in knowledge and in capabilities. Leaders of generating companies, grid companies, and consuming companies often do not understand the business case for sharing data in an open fashion. Some believe that benchmarking efforts will do little more than help their poorly performing peers. Similarly, consumers are often unaware of how much they might save through conservation and retailers must figure out how different types of consumers can be motivated to participate in energy-saving schemes. Many companies lack experts in data analytics, and in many organizations IT departments are not in a position to take on the additional capabilities and responsibilities required to carry out open data initiatives.

IMPLICATIONS FOR STAKEHOLDERS

Open data can help stimulate nearly $500 billion a year in value across the electricity ecosystem, while saving fuel and reducing greenhouse gas emissions. But turning this opportunity to reality will require effort by all stakeholders.

- **Electricity users.** Users of electricity, including residential, commercial, and public-sector users, have the most to gain from the use of open data applications. But they need to understand the potential benefits and how to capture them. They also need to educate themselves about associated risks. Users can take advantage of open data and open data-enabled applications and services to optimize their investments in energy efficiency, and manage their overall consumption, by participating in MyData demand-management initiatives or in industry benchmarking efforts. They can lobby utilities and regulators to open more consumption data and engage in public discussion about energy-efficiency programs and provide feedback about the information that is presented to them. To ensure that their privacy and confidentiality concerns are understood, electricity users can work with governments and energy firms to shape how energy data are released and used.

- **Government.** There are many ways in which policy makers and regulators could enhance the value captured through the use of open data in the electricity domain: creating incentives to capture value from open data, leading the educational effort that will be necessary to bring companies and customers on board, acting as energy user role models, and putting in place sensible safeguards and standards.

  Laws and regulations can encourage the capture of value through open data levers, and regulators can realign currently misplaced incentives. To motivate investment in energy-efficiency technologies, governments could consider adjusting rate-capped utility regulations, for example.

  As leaders in educating other stakeholders about the benefits and risks of open data, lawmakers and policy makers can conduct public awareness campaigns, particularly about energy-saving opportunities, and they can make the economic case. Policy makers will also need to do some research to understand what motivates different types of consumers to save energy and what incentives could be effective.

  As gatherers and distributors of open data, governments can influence creation of standards for data reporting and data formats, as well as user-friendly websites. Furthermore, governments should consider the benefits of aggregating and making open various sets of data, such as reported electricity consumption data. Governments can also encourage entrepreneurs to create applications and communication tools to help inform consumers, citizens, and producers. These might include carbon footprint calculators for various power generators, permit approval models. Officials will also need to pay attention to rapidly evolving technology and operating models, especially in power generation, to ensure that aggregated data, best practices, and benchmarks can be shared with as many parties as possible

  As a large consumer of electricity, governments can act as role models for their constituents by serving as lead users and examples for using open data to inform energy decisions.

When appropriate, policy makers can provide guidance on uses of open data. This could involve discussions on what types of data or data usage might compromise individual privacy, company proprietary data, and national security. Topics to be considered include how consumer open data might be used (as in advertising), corporate confidentiality, and data security.

- **Companies.** Electric companies and independent data providers have much to gain from open data. They can establish clear, easy-to-understand standards, metrics, and interfaces to facilitate data sharing across the industry. They also can take responsibility for safeguarding their customers' data. They can establish industry standards for collection, storage, and use of consumption information. Companies should clearly communicate these practices to customers. In addition, they can promote participation in consumer open-data systems by offering easy-to-use applications and publishing consumption reports that provide straightforward recommendations for improving efficiency. These self-serve efforts can be backed up with knowledgeable customer service.

  Electricity companies have a unique opportunity to release open data that enables third parties to build innovative products and services around open-data initiatives. This might include customer applications that "gamify" the practice of energy conservation. In their own operations, electric companies can capture value from open data by using and sharing open data about operations. To do so, they might have to fill gaps in talent and capabilities.

  Further value potential lies in providing transparency about the environmental impact of electricity generation. Where consumers have choice, they can use this additional information when selecting an electricity provider. Third parties can also compile such information. Third parties may be particularly important, since they can be a trusted neutral repository of information that companies are reluctant to share. Data aggregators should work with the producer companies to set guidelines on levels of detail of information and ways in which to protect the identity of companies and customers, as well as protocols for safeguarding the security of raw data.

5. Open data in oil and gas

Despite the evolution of alternative energy sources, much of the global economy still runs on oil and gas and will continue to do so. With rapid growth of developing economies, demand for these fuels is expected to grow 56 percent, from 524 quadrillion BTUs a year in 2010 to 820 quadrillion BTUs in 2040.[53] Filling that demand requires increasingly large investments in exploration and production to find reserves in hard-to-reach places (e.g., ultra-deep-water drilling) and to develop fields using new technologies such as hydraulic fracturing, which can release oil and gas trapped in shale formations.[54] The cost of developing reserves has risen steadily over the past decade as exploration has moved toward increasingly complicated basins.

Open data has the potential for significant impact in this technology-dependent sector. Exploration and production companies have already amassed immense quantities of seismic data to pinpoint the most likely reservoirs. Companies also use real-time data to guide production operations in the field. The introduction of open data techniques, such as sharing data about new fields and production performance metrics, can help companies raise the productivity of investments in exploration and production, as well as improve the efficiency of operations. Open data can also expose variation and encourage standardization and collaboration in a complex global industry with myriad geopolitical interests, regulatory mechanisms, and economic systems.

In this chapter we examine how open data and data sharing can be used across the oil and gas value chain, from capital-intensive upstream operations (exploration, development, production) to midstream activities (transport, trading, processing of natural gas, refining), and downstream operations (marketing, retailing, consumption). We identify five levers for using open data across the value chain:

- **Optimized upstream investment.** Sharing geological and seismic data as well as regulatory filings can lead to increased discovery. Capital productivity benchmarks based on shared data can also reduce investment costs and avoid construction delays.

- **Efficient upstream operations.** The efficiency of exploration and production can be increased by identifying specific opportunities through comparing internal performance against operational benchmarks.

- **Optimized midstream and downstream investment.** Publicly available geological data as well as projections of oil and gas demand, deal flow information, economic trends, and demographic data can help companies make better midstream and downstream investments, such as choosing plant locations to serve the most promising growth markets. Sharing benchmarking

53 *International energy outlook*, Energy Information Administration, May 2013.

54 For more on advanced oil and gas technologies, see *Disruptive technologies: Advances that will transform life, business, and the global economy*, McKinsey Global Institute, May 2013.

and best practice data can also help reduce inefficiencies in the construction of infrastructure.

- **Efficient midstream and downstream operations.** The efficiency and reliability of refining, processing, and retail operations can be improved through better supply and demand management, combining traditional forecasting and planning methods with shared data sets (for example, demographic shifts, water availability projections, and industrial control data). Operating benchmarks can create additional value.

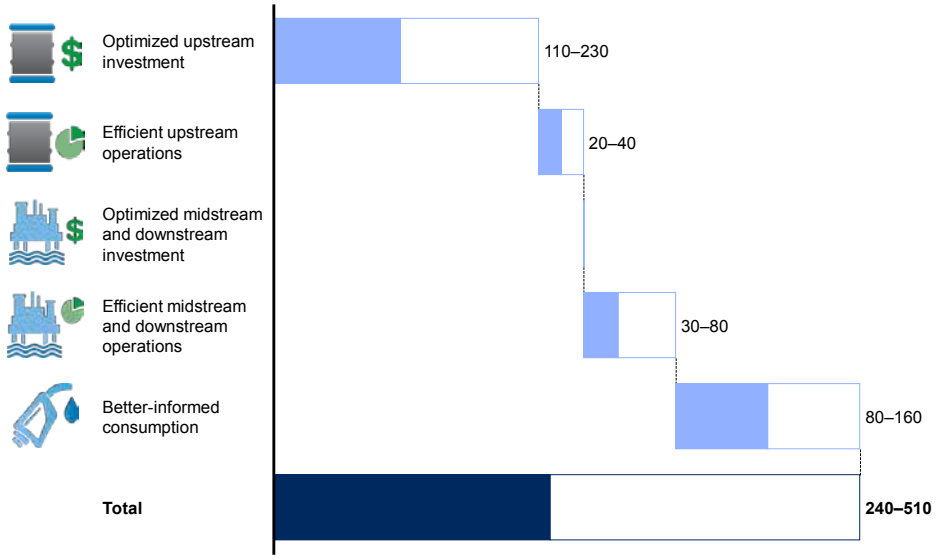
- **Better-informed consumption.** Open data can help both residential and commercial energy users control their consumption. With data showing their own usage in comparison with that of similar customers, users can make informed decisions about consumption and take steps such as investing in more efficient heating systems and insulation.

We estimate that across the oil and gas ecosystem, the use of open data can help to unlock $240 billion to $510 billion in value per year (Exhibit 9). Companies stand to gain from improving the returns on investment in exploration and development, as well as from improvements in operations. Energy users stand to benefit through increased transparency about use, which helps them reduce consumption through energy efficiency.

Exhibit 9

Five levers can help oil and gas industries realize $240 billion to $510 billion in value per year

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

There are many barriers to realizing the value potential in the use of open data in oil and gas. The most significant is demonstrating that it is in the mutual interest of competitors to share certain types of data, which have historically been regarded as sensitive (e.g., seismic data). Also, there is wide variation in how governments in different jurisdictions collect and share relevant data, such as lease records—often data are not collected or shared. Even when made available,

data might not be provided in machine-readable form and must be gathered manually. This lack of standardization limits comparability and the number of insights that can be drawn. Finally, issues of energy independence and market dynamics require governments to weigh the policy implications of the data they require companies to report and the data that they make open to the public.

A large amount of data is collected and used in the oil and gas industry. The main sources of data are oil and gas companies, but governments, third-party data providers, interest groups, and consumers also have valuable data that can be used in open data applications (see Box 9, "Sources of open data in oil and gas"). Additionally, governments have the ability to influence the amount of data that oil and gas companies share via regulation and legislation.

Box 9. Sources of open data in oil and gas

- **Companies.** Oil and gas companies collect performance data in two main categories: capital expenditures and operations. Upstream investment data include exploration data—4D seismic readings (describing three-dimensional data over time) from existing and potential sites. Upstream and midstream operating data include metrics such as platform operation metrics and pipeline costs, as well as an immense amount of sensor data collected in production fields. Downstream data include costs and other operational metrics from refining processes. Companies also gather information for health, safety, and environmental reporting that is required by regulators, and they gather industry data: competitive intelligence, data on deal flows, and aggregate data on how oil and gas are transported.

- **Governments.** Oil and gas is a regulated industry that plays an important role in the economic and geopolitical decisions of many nations. As a result, governments often collect a significant amount of data: permits to drill and equipment to be used, as well as health, safety, and environmental data. Some governments also collect performance data, such as discoveries, output of fields, and revenue. By aggregating firm-specific data, governments are able to provide macroeconomic views of energy markets. Some governments share supply, demand, and pricing projections. Other government data, such as population figures, statistics on crop yields, economic reports, patent filings, traffic data, international trade data, and weather data, are also relevant to the oil and gas sector.

- **Consumers.** Through their purchasing, oil and gas customers generate valuable data regarding consumption of fuel for residential, commercial, and industrial uses. Customers may also contribute data to various crowdsourced efforts, such as gasoline price comparison apps that rely on point-of-sale data contributed by consumers using smartphones.

- **Third-party data brokers.** Independent data providers serve as neutral third parties that can be trusted to collect firm-level performance data, analyze and share it with the industry in aggregated and anonymous reports. Brokers collect data across multiple geographies and technologies to publish benchmarks and comparative data. Often, this involves standardizing and aggregating data. Third parties may also collect data about retailing, such as pricing data.

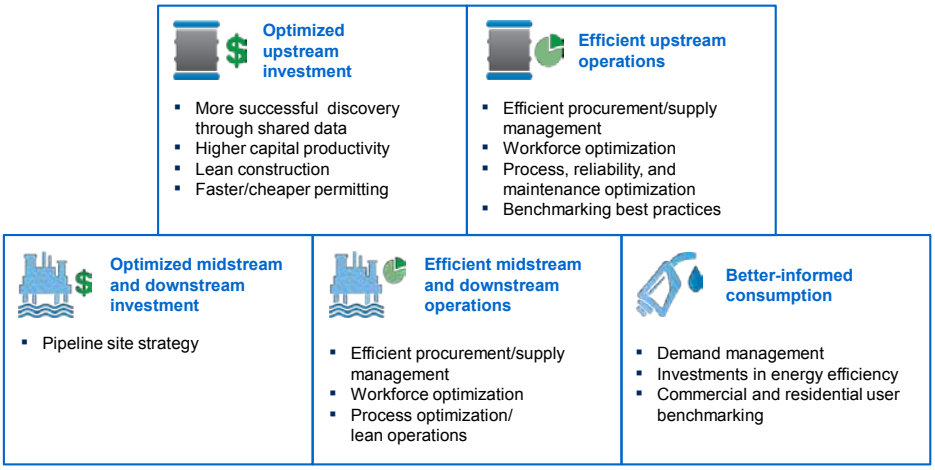
- **Other organizations.** Trade associations, non-governmental organizations, and intergovernmental organizations also collect information about oil and gas industry performance and practices. The International Association of Oil and Gas Producers provides a forum for member firms to share best practices, mostly on health, safety, and the environmental performance. Other intergovernmental organizations such as the International Energy Agency produce reports on global oil and gas markets.

HOW FIVE LEVERS CREATE VALUE IN OIL AND GAS

We estimate that the use of open data could help to unlock $240 billion to $510 billion annually in the oil and gas domain. Half of this value could arise in upstream exploration and production, where increased collaboration and sharing of data, including best practices and benchmarks, could spur increases in efficiency and output. However, a large portion of that value may not be easily unlocked, since it would require sharing data and methodologies among many parties, including direct competitors that view such information as a source of competitive advantage. The remainder of the value lies in efficiency improvements, for both oil and gas companies and their customers. For each value lever, we identify several possible approaches (Exhibit 10).

Exhibit 10

How five levers create value in oil and gas



SOURCE: McKinsey Global Institute analysis

Optimized upstream investment

Open data has the potential to unlock significant value in exploration and production functions in two ways. First, greater sharing of seismic data and information about drilling methodologies can increase the success of exploration activities, leading to improvements in exploration metrics such as reserve replacement ratios (RRRs) and costs per barrel.[55] Second, sharing construction, procurement, and other benchmarks and best practices can help to deliver costly capital projects more efficiently by avoiding cost and schedule overruns.

For the past decade, upstream activity has been characterized by rising exploration costs, declines in reserve replacement ratios, and longer time to production, especially for major oil and gas companies, Exploration, by definition, is a speculative activity, in which success often depends on making good decisions with incomplete information. Blocks of land are drilled without positive results, and then sometimes subsequently re-explored at a slightly different location or depth with success. This occurred in the North Sea as recently as 2012, when Statoil discovered 140 million to 270 million barrels of recoverable oil in the same blocks that had been drilled by Elf in 1976. The exploration had missed the discovery by 25 meters, and for the next 36 years, the entire Utsira

55 The replacement ratio is defined as the ratio of discovered oil and gas reserves compared with the amount of oil and gas produced.

High region had been largely written off by the industry because of its perceived complex geology and Elf's unsuccessful exploration.[56]

Since the 1970s, advances in technology have made it possible to explore deeper basins, more subtle traps, and complicated reservoirs that range from onshore and shallow water to deep and ultra-deep-water sites.[57] But as the possibilities have expanded, the complexity has increased, raising costs and depressing success rates. Given these challenges and capital constraints, companies do not have the resources to pursue all of the prospects identified for exploration. Some experts estimate that the ratio of prospects to explored targets is as high as 50:1.

Open data and data sharing can raise the rate of successful discovery. Increased sharing of information across the industry, possibly with government intervention, could help companies develop more complete assessments of what they might find before committing hundreds of millions of dollars to a new exploration project. Sharing seismic, permitting, and other data across companies could lead to more optimal investments and thereby higher production by, for example, creating better understanding of the geology across an entire basin. Global sharing of all seismic data is unlikely to happen, but increased sharing, even among a few players, does have the potential to provide value to the market that is not currently captured. And sharing R&D findings about technology that benefits all players, such as developments in blow-out prevention technology, can also reduce costs and environmental impact, while increasing safety.

Exploration and production firms could also use crowdsourcing to improve their performance. In gold exploration, the Canadian mining company Goldcorp published 400 megabytes of geological survey data about its Red Lake, Ontario, property in 2000. Goldcorp offered $575,000 in prizes for public submissions identifying likely locations of gold within the mine. At the time, Goldcorp's geologists were concerned about the release of data that were considered highly proprietary. However, after collecting submissions from more than 1,400 participants from more than 50 countries, the results yielded 110 targets, of which 80 percent were productive. This has led to the extraction of more than $3 billion in gold.[58] Crowdsourcing has been used in a limited way in oil and gas already. In 2011, after the Deepwater Horizon spill, the X Prize Foundation conducted the Wendy Schmidt Oil Cleanup X Challenge, which was sponsored by Shell, to spur innovation in spill cleanup technologies.[59]

Governments, which have economic and national security interests in finding energy supplies, could take the lead in data-sharing initiatives such as requiring seismic data to be submitted in exchange for licenses to explore. In the United Kingdom, the Department of Energy and Climate Change runs the Fallow Initiative: blocks and discoveries are declared "fallow" if the companies with licenses are not progressing with exploration and production and their blocks and discoveries are then made available to other parties. Nearly 200 blocks have been relicensed since the program began in 2002. Similar legislation was proposed in the United States in 2008. While neither program releases data about the blocks that are up for re-use, one could envision a regulatory scheme where licenses are

56 Jeremy Beckman, "Statoil aims to squeeze more out of Oseberg," *Offshore*, October 1, 2012.

57 Geological features that contain oil and gas, preventing leakage or escape into the surrounding environment.

58 Linda Tischler, "He struck gold on the Net (really)," *Fast Company*, June 2002.

59 The challenge's website: www.iprizecleanoceans.org.

initially granted only with submission of seismic data, and those data are made open, along with the original licensing applications, when blocks are put up for re-use.

While there are challenges in sharing geological data that many companies consider to be highly proprietary, it could be easier to reap the benefits of open data in exploration and production operations. By sharing benchmarked metrics through a neutral party, oil companies can identify areas of potential savings and make their capital investments more productive. Through careful application of benchmarks to identify areas for improvement and the implementation of best practices for managing project costs and timetables, we estimate that companies can save 15 to 25 percent per project.

Additional incremental value can be gained by making more siting and permitting data held by governments machine readable and accessible. In the United States, data on mineral rights negotiations are not kept in a standard digital format and formats vary from county to county. The resulting inability to quickly identify mineral rights has given rise to an entire profession of land managers who conduct due diligence by going county by county, courthouse by courthouse, to confirm rights on land that oil companies are considering as drilling targets. If these data were to be standardized and digitized, the oil and gas industry could improve the efficiency of its land management activities.

Across upstream exploration and production capital projects, we estimate that the use of open data could lead to $110 billion to $230 billion of value annually. This includes an estimated 10 to 25 percent savings in exploration capital expenditures and improved performance in exploration activity.[60]

Efficient upstream operations

Oil and gas exploration and production companies have much to gain from sharing data that can improve the efficiency of operations after the discovery is confirmed, the field is developed, and the process of extraction begins. This type of data sharing is already seen in upstream operations and is an excellent target for open data initiatives aimed at reducing costs, limiting environmental impact, and increasing safety. Benchmarking to improve operations can be done across geographies, but is most commonly conducted today within a single basin, such as the North Sea. Neutral third parties can aggregate historical data and share various production, cost, and safety performance metrics to help operators highlight where they have opportunities to improve.

While data are often collected by individual firms, they are not always shared or available in standardized formats. Of course, sharing these metrics is not enough—operators must use this information along with data on best practices to devise practical steps for realizing the value from safer, more efficient operations. After identifying opportunities by comparing their performance with shared metrics, operators can devise practical steps to realize the value from safer, more efficient operations.

We estimate that through insights gained from sharing of benchmarks, and the subsequent application of best practices in upstream operations, oil and gas

60 This assumes that finding costs of the bottom 50 percent of explorers could be brought to the median cost level and reserve replacement ratios are increased through more data sharing.

companies could save $20 billion to $40 billion annually in upstream operations, roughly 10 to 20 percent of global annual operating expenditures.

Optimized midstream and downstream investment

Midstream and downstream investments include building oil and gas pipelines, as well as processing plants and refineries. Open data has the potential to optimize investments in this area, through improved links between supply and demand, more efficient siting and permitting processes, and sharing of capital productivity benchmarks.

Improved information on the supply and demand for fuel can lead to more efficient transport operations, by informing optimal pipeline siting decisions. Transportation of fuel through pipelines is several times less expensive than moving gas and oil with trains or trucks.[61] However, pipelines have very steep upfront costs, making pipeline companies reliant on sustained demand to recover the investment costs and remain economically viable. If open data about the transportation of all crude oil, refined products, and natural gas were available, pipeline companies could optimize siting decisions.

It is also possible to use capital productivity benchmarks to find performance improvement opportunities in construction and other areas. We did not estimate the total size of the opportunity, because the total size of midstream and downstream capital expenditures is relatively small (about one-tenth of upstream investment), and the field is relatively mature.

Efficient midstream and downstream operations

As in upstream operations, the potential for value creation with open data in midstream and downstream operations lies in sharing benchmarks that can help refiners identify opportunities to improve safety, reliability, and efficiency. The impact of these operational improvements can be significant. For example, due to the maturity of the industry, refiners vary little in their technology and processes. Yet top-quartile performers still earn significantly higher returns than bottom-quartile performers, an important advantage in a business with single-digit EBITDA margins. Indeed, operational efficiency is the single biggest differentiator among refiners.

Third-party data aggregators can be the source of sharable operating data and benchmarks. This information can be used in lean operating initiatives, best practice planning, and benchmarks for maintenance. We estimate that refineries globally have the potential to unlock $30 billion to $80 billion annually in value through savings that are catalyzed by the use of this sort of open data, roughly $1 to $2 savings in refinery costs per barrel.

Better-informed consumption

The primary source of value from open data for consumers of oil and gas would come from reduced consumption and lower fuel costs. The same demand management techniques based on open data that apply to electricity use apply in oil and gas consumption. A detailed discussion of those measures can be found in the preceding chapter on page 70.

61 Matthew Philips and Asjylyn Loder, “Amid US oil boom, railroads are beating pipelines in crude transport,” *BloombergBusinessweek*, June 13, 2013.

We estimate that $80 billion to $160 billion of value can accrue annually to consumers through savings enabled by open data. This is approximately a 10 to 20 percent reduction in building energy consumption, a 50 to 66 percent reduction in fuel costs through more prudent purchases of vehicles, and corresponding reductions in emissions.

BARRIERS AND RISKS

Implementing open data in the oil and gas sector will require overcoming significant barriers, including persuading some of the world's largest corporations and most intense competitors to share information. Other barriers involve technical issues and a lack of knowledge about the benefits of sharing information.

The largest barriers undoubtedly are behavioral. Companies across the value chain are concerned about the improper use of their proprietary data. Even though data are sometimes shared by field operators or gathered and controlled by independent third-party data brokers, companies worry that specific performance metrics will be identified as theirs—showing competitors how well or poorly they are performing—or accidentally revealing proprietary information. Companies are even more protective of exploration data, especially in areas where land rights have not yet been secured.

There are also technical barriers, including a lack of standard formats and variation in data filed with regulators and public agencies. Because different jurisdictions require different data and reports, it can be challenging to come up with comparable metrics and draw useful conclusions from aggregated data. Similarly, permits are not always machine readable and might be available only through in-person inspection. Environmental impact reporting is another area where apples-to-apples comparisons are difficult because of a lack of reporting standards. "Noise" in the data can be a challenge as well.

Sometimes, the necessary data have not yet been made openly available. In Europe, under the Energy Performance of Buildings Directive, energy consumption data about many buildings have been made available. In New York City, the government has released a similar data set on the energy use of commercial and multifamily residential properties. These efforts allow owners and tenants to compare their energy use with that of similar businesses or buildings. However, there are many places where such data are not available, making it more difficult for landlords and tenants to make decisions such as identifying buildings that are in need of retrofitting (e.g., with new HVAC equipment, windows, and insulation).

As in other applications of open data, consumer concerns are a major obstacle. Consumers who use oil or gas to heat their homes can participate in demand-management programs that depend on capturing hour-by-hour data on natural gas use—through MyData reports or through in-home displays. This information provides the basis for customer actions such as lowering thermostats and investing in insulation. However, some consumers fear that the data could signal to criminals when the home is empty. Similarly, in the corporate world, if a rival can detect a surge in energy use, a company's accelerated production plan might become known.

IMPLICATIONS FOR STAKEHOLDERS

Across the oil and gas universe, from independent exploration companies to motorists at the pump, there is value to be gained from the use of open data. Here we discuss the implications for various stakeholders.

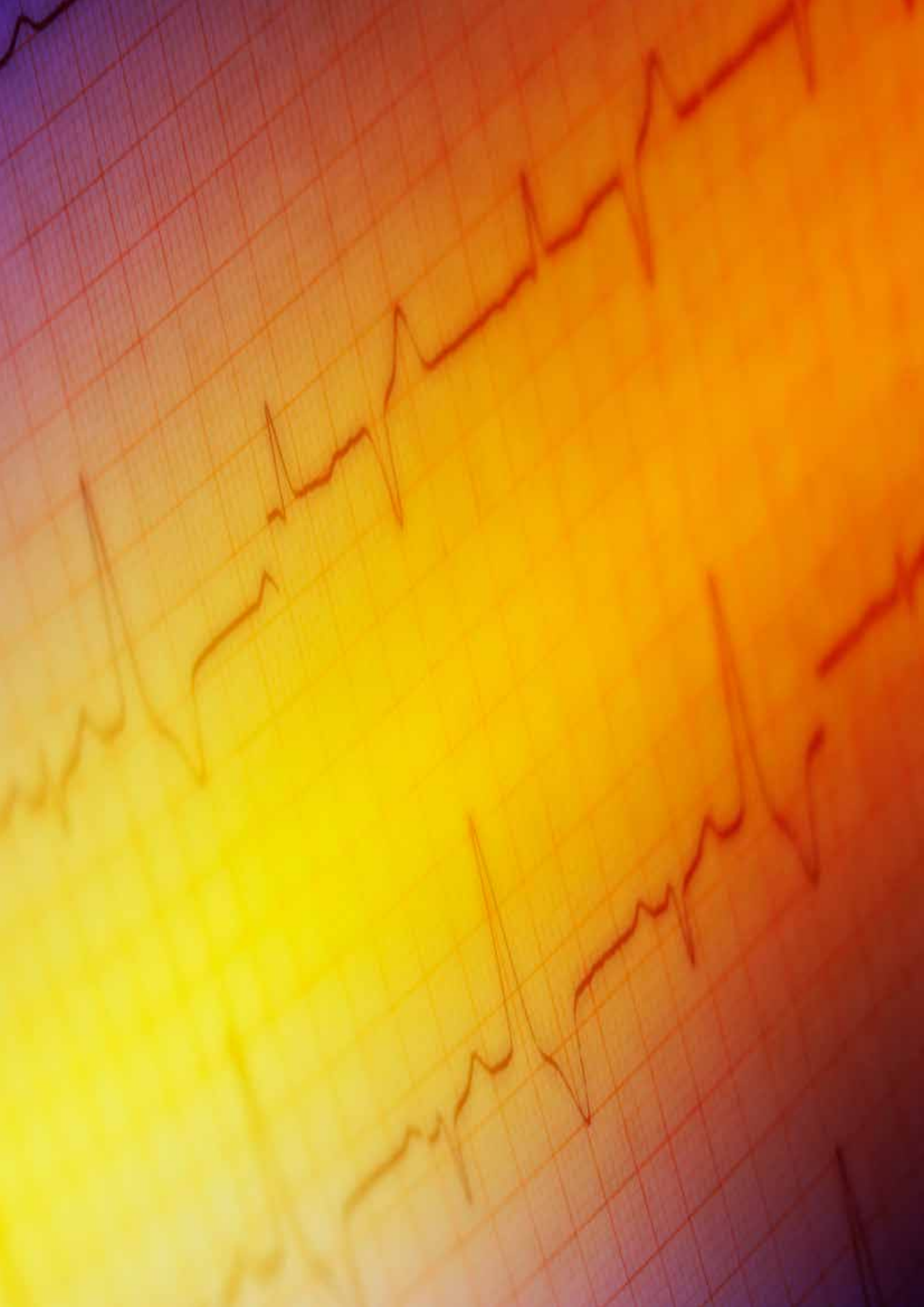
- **Oil and gas producers** have a great deal to gain from the use of open data to improve the productivity of their investments and the efficiency of operations. These improvements can help them find the sources of oil and gas to meet future demand and bring down the costs of developing and operating in new basins. In addition to investing the time and effort to understanding the business case for collaborating in open data initiatives, companies can work collaboratively and with third parties to establish standardized formats for reporting data as well as safeguards (including contractual requirements) for maintaining the security and anonymity of data that are contributed to industry-wide open databases.

- **Gasoline and natural gas retailers** will need to take steps to assure consumers and business customers that data about their usage will be safeguarded as retailers apply open data techniques to customer-facing operations. Oil and gas companies can create easy-to-understand reports of consumer energy use with specific recommendations for savings and create online resources for easy access to company expertise.

- **Policy makers and regulators** have wide-ranging interests in seeing that open data is used effectively to improve their nations' energy supplies, ensure compliance with environmental standards, and bring benefits to energy users. They can encourage or require the development and use of standards. They can also foster an environment for entrepreneurs to create applications and communication tools using open data to better inform consumers, citizens, and producers (e.g., showing the carbon footprint of various petroleum sources, or modeling the likelihood of permit approvals).

  Policy makers can start by convening various stakeholders, including community groups and commercial building owners, to establish guidelines for the responsible release of consumption data, taking into account privacy and confidentiality concerns.

- **Consumers** would need to participate in energy monitoring and conservation efforts in order to gain the value available through use of open data in energy management. This may require an effort to learn how to use new applications, read energy reports, or provide feedback to help make these reports and applications more usable. To make sure that their energy use information is not used in ways they don't want, customers (individuals and commercial users) may need to work with government and utilities to shape policies and practices.



6. Open data in health care

The use of data to improve the quality of health care and make delivery more efficient is reaching a tipping point, with open and proprietary data playing a leading role. The amount of data being collected, analyzed, and shared among health-care stakeholders has reached critical mass, with growing volumes of digitized medical records, aggregated research and development data, and data that government has collected over the years. These big data sets are yielding critical insights into what are the most effective therapies for specific types of patients, enabling hospitals to isolate common causes of costly hospital re-admissions and allowing insurers and other payers to identify variations in care delivery that add needless costs. In addition, individual consumers are finding new ways to manage their health and fitness through MyData, which is made available by care providers or generated by consumers by using exercise monitors and other devices.

We identify five levers for capturing value in health care through the use of open and proprietary data. They span the full range of health-care stakeholders from patients to research scientists. Based on previously published McKinsey research, we focus on levers that can be applied in the United States, but many of the levers that we identify work in other nations and have been applied in health-care systems around the world.[62]

- **Right living.** Using data to make better lifestyle decisions and to manage treatments. With access to MyData and self-generated data (from exercise monitors, for example), patients can prevent lifestyle-related illnesses and comply more successfully with instructions for taking medication or adjusting diet and exercise routines.

- **Right care.** Using data analytics on data from thousands (or millions) of patients to discover which treatments have been most effective and using these insights to ensure that patients get the most timely and appropriate treatments.

- **Right provider.** Using data analytics to match the best sources of treatment—facility, physician, nursing staff—for a particular patient and condition. Based on proven outcomes, patients and payers can specify what facility and caregiver is best in every case.

- **Right value.** Using data to ensure the cost-effectiveness of care. This includes providing consumers with information to make informed choices about care, based on cost as well as on outcomes. It also includes use of open data to detect fraud.

62 This chapter relies heavily on research documented in *The "big data" revolution in health care: Accelerating value and innovation*, McKinsey Center for US Health System Reform and McKinsey's Business Technology Office, January 2013.

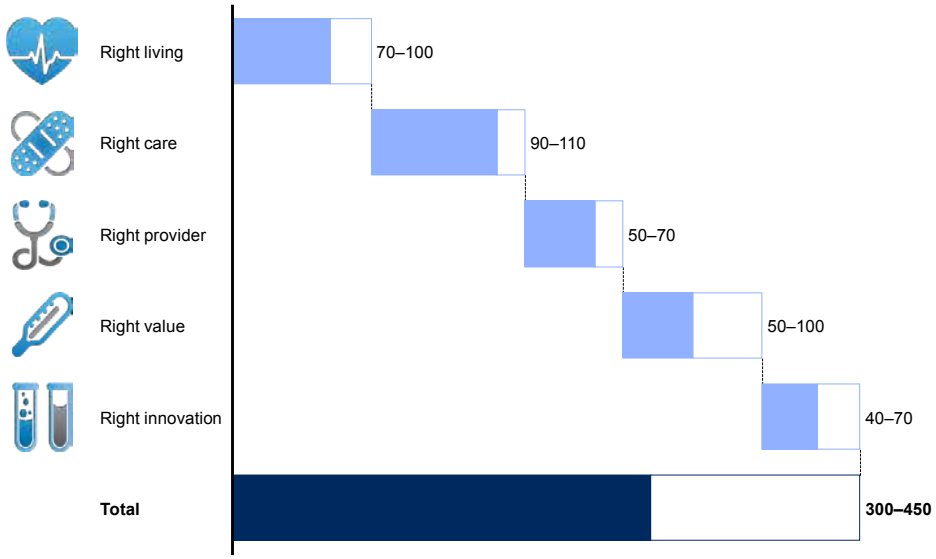
- **Right innovation.** Raising the productivity of R&D by makers of drugs and medical devices sharing data (where possible), using open data to find patients for trials, speeding products to market, and spotting potential problems once products are in use.

We estimate the use of open and proprietary data in health care could help generate value of $300 billion to $450 billion per year in the United States (Exhibit 11). Most of this value comes in the form of cost savings to providers, payers, and patients.

Exhibit 11

Five levers could reduce US health-care costs by $300 billion to $450 billion per year

$ billion



NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis

Several changes would have to be made in the US health-care system for the full value of open data to be realized (many of these changes would accelerate value creation in other countries as well). The most fundamental change would be to shift medicine and caregiving to data-driven approaches, where physician decisions about treatment would be informed by results from thousands of patients. Payment systems would also need to be adapted since conventional means of controlling costs, such as negotiating prices of per-procedure fees, are not geared to taking advantage of the insights that open data provides.

There are also technical and organizational barriers, including the inability of many health-care data systems to provide standardized data; even within one pharmaceutical firm, shareable data often remain siloed. Finally, there are concerns about privacy and confidentiality—the consequences of mishandled medical data can be extremely serious. Government, providers and payers will need to make sure that effective systems are in place to keep shared medical records confidential.

As noted, the health-care industry has no shortage of raw data. Information about the health of patients, diseases and injuries, treatments used, and fees charged is collected at every stage of care and in almost every setting (see Box 10, "Sources of open data in health care"). As EMRs proliferate, data from clinical notes, laboratory values, and physical exam findings will also be available.

Box 10. Sources of open data in health care

- **Providers.** The amount of patient clinical data captured by health-care providers has grown exponentially. In 2005, only about 30 percent of US physicians and hospitals used even basic electronic medical records (EMRs). By 2012, 50 percent of physicians and nearly 75 percent of hospitals were using electronic records. Furthermore, around 45 percent of US hospitals are now either participating in local or regional health-information exchanges or are planning to do so in the near future.

- **Payers.** Insurance companies and government health plans hold claims and cost data that describe what services were provided and how they were reimbursed.

- **Pharmaceutical and medical products manufacturers.** Drug and medical device manufacturers generate massive amounts of clinical data as they test new products.

- **Consumers.** Increasingly, consumers are generating their own health data, by using connected devices such as running shoes and wristbands that monitor exercise, physical activity, or sleep. Networked scales, pill bottles, and blood glucose monitors also collect and share clinical information with providers. Consumers also share health-care data online, through social media sites and on forums focused on specific diseases such as PatientsLikeMe.

- **Government.** The US Department of Health and Human Services (HHS) releases data from agencies such as the Centers for Medicare and Medicaid Services (CMS), the Food and Drug Administration, and the Centers for Disease Control and Prevention. Examples include the Healthcare Cost and Utilization Project (HCUP, from the Agency for Healthcare Research and Quality, AHRQ), which is the largest all payer data publicly available; the Adverse Event Reporting System (AERS), which is the database that supports the FDA's post-marketing safety surveillance program; and International HapMap, which is a multinational effort to compare genetic sequences of different individuals to find genes that affect health, disease, and individual responses to medications and environmental factors.

  The Affordable Care Act, enacted in March 2010, includes a provision that authorizes HHS to release data that can increase transparency in the markets for health care and health insurance. In the United Kingdom, data made public through the National Health Services (NHS) Choices program allows patients to compare hospital ratings and to review qualitative reviews of doctors.

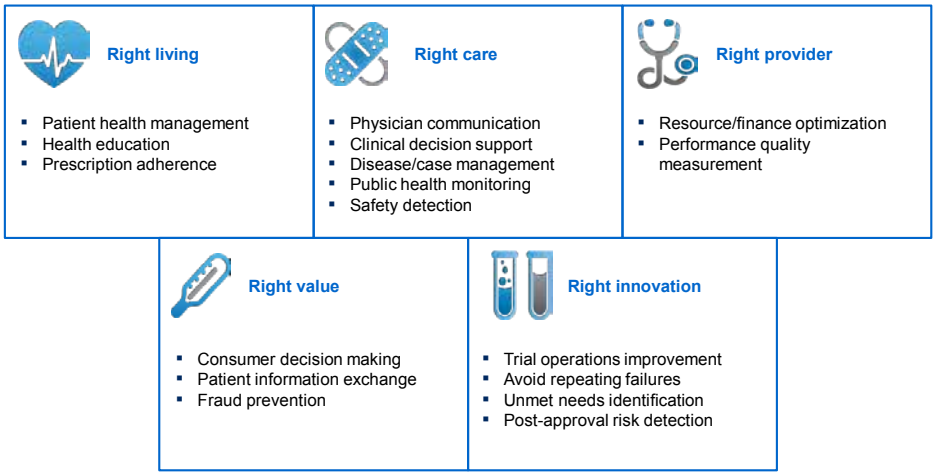
- **Third-party data brokers.** These companies capture patient behavior and sentiment data that describe patient activities and preferences that can be relevant to health care, from social media to eating habits. These developments allow stakeholders access to a broader range of information.

HOW FIVE LEVERS CREATE VALUE IN HEALTH CARE

We have identified five key levers to create health-care value with the help of open data. Value is derived from optimizing both health-care spending (cost) and patient impact (outcomes). Each lever can be applied in a variety of ways (Exhibit 12).

Exhibit 12

How 5 levers in health care can create value



NOTE: Applications fitting in multiple customer categories were counted multiple times.
SOURCE: McKinsey Global Institute analysis

Right living

An enormous and largely untapped source of value in health care lies in the use of data by patients to manage their health to avoid illness and to get better results from treatment when they are ill or suffer from a chronic condition. The major focus is providing patients with the information to make the best lifestyle choices and manage their treatments effectively.

Healthy consumers can reduce the chances of lifestyle-linked conditions and illnesses such as hypertension and diabetes by adjusting diet and exercise. Health-care data that are made more liquid by consumers can help in this effort by enabling health-care providers or applications to identify at-risk individuals. This is done by combining information collected about patients such as exercise habits with demographics information and analyzing outcomes across different patient populations. At-risk patients can then be targeted for health education or for assistance in preventing illness—for example, by recommending screenings or issuing personal reminders.

One of the greatest potential impacts of open data is assisting patients with drug adherence. It is estimated that failing to use medication as prescribed for chronic conditions costs $100 billion to $289 billion a year to the US health-care system.[63] For a variety of reasons, patients fail to take their medications as directed, often leading to hospital re-admissions, medical emergencies, and complications. More liquid data can be used to improve patient adherence. For example, applications that use open data through social media can be used by patients' friends and

63 Meera Viswanathan et al., "Interventions to improve adherence to self-administered medications for chronic diseases in the United States: A systematic review," *Annals of Internal Medicine*, volume 157, issue 11, December 4, 2012.

families to encourage adherence. Also, data shared by individual patients or gathered automatically (by smart pill dispensers, for example) can be used to remind patients to use therapies as prescribed.

A startup company called Ginger.io has developed a smartphone app that can detect when a patient is not using drugs correctly or has changed behavior in a way that indicates a potential medical emergency. The app collects data about motion and activities to establish a base pattern and then looks for deviations that might signal trouble; a sharp drop in activity, for example, might indicate that a patient has stopped taking antidepressants or that a patient has fallen. Irregular sleep patterns could signal that an anxiety attack is imminent.

We estimate that right living techniques can enable $70 billion to $100 billion annually in value through reduced care costs by preventing disease and raising adherence rates.

Right care

The best health-care outcomes are based on getting the most timely, appropriate treatment available and preventing avoidable mistakes, such as administering a medicine to which the patient is allergic. Right care requires a coordinated approach across care settings and providers; all caregivers must have timely access to the same information and work toward the same goal to avoid duplication of effort and sub-optimal strategies.

Research shows that sub-optimal care is often the result of poor communication between patients and doctors and among doctors about a patient's medical history, current medications, allergies, and other relevant data. This leads to inappropriate or redundant care that can result in complications and raise costs. Shareable electronic medical records are helping to avoid these problems by keeping a single consolidated record for each patient in a file that all caregivers can access. In the United States, patients can gain access to their own medical records from participating providers—their medical MyData—through an initiative called Blue Button, which was pioneered by the Department of Veterans Affairs. In France, patients can carry their entire medical histories on a smart card—a card with a memory chip that any doctor or care facility can use. The system can also flag possible drug interaction issues for a specific patient before a physician selects a drug treatment or radiologic test. The data needed for such systems are not always shared across systems today, but it is likely that in the next decade more data will be integrated from multiple sources to make these systems more robust.

In addition to heading off known adverse drug interactions, open data can be used to detect associations between certain treatments and adverse effects. For such programs to work, health-care systems will have to pool information on health outcomes, claims, and other factors in real time. Potential sources for these data include payers, providers, and government agencies, such as the US Centers for Medicare and Medicaid Services.

Another way in which liquid data can be used to improve care is to mine data to find patients who can benefit from specific treatments. Doctors can contact their patients who have certain medical and acute conditions to provide information about the diseases and offer nursing support to manage their conditions. Advanced analytics of claims history, demographics, and other factors can be used to identify patients who are most likely to benefit from extra assistance.

A final aspect of "right care" is mounting a vigorous response to disease outbreak. Public health agencies collect data from emergency rooms and other sources to detect outbreaks of certain diseases so that people can take steps to protect themselves. Google has shown that it can map a flu outbreak in real time by monitoring the number and locations of users searching for flu and related topics. A startup called Propeller Health has created a GPS-enabled tracker that monitors inhalers. It can be used to detect when environmental conditions such as high pollen counts might trigger severe asthma problems and can alert patients to take precautions.

We estimate that open and proprietary data can help enable $90 billion to $110 billion a year in value by reducing cost of care, using proven, evidence-based care pathways, and coordinating care among multiple physicians, clinics, and payers across the system.

Right provider

"Right provider" means finding the right match of provider skills for the task—identifying the doctors and settings that provide high quality of care and determining what resources are needed to deliver care (for example, determining which procedures are done best by a physician assistant or nurse rather than by a doctor). Using health-care outcomes data, patients and payers can determine the best choices of providers. Within institutions, the same approaches can be used to optimize operations; hospitals can use outcomes, benchmarks, and other shared information to help identify places where improvements can be made.

Optimizing provider decisions depends on access to performance data, which must be made accessible across organizations. Performance data can be used to align incentives with outcomes (paying doctors for successful treatment and efficient care, for example, rather than for procedures). This can raise the quality of care and give patients better options. For example, in the US state of Arkansas, all Medicaid providers who treat upper respiratory infections, pregnancy, attention deficit hyperactivity disorder, and other select diseases receive a report on the cost and quality of treatment. The aggregated data on all of a provider's claims help doctors understand how their practices compare with other providers in the state, giving them a way to identify opportunities to improve. The US state of New York and the United Kingdom's National Health Service have opened up performance data that they collect about hospital performance and different types of care. These data can be used by patients, doctors, payers, and the public to identify the best care available in their communities.

We estimate that right provider measures could enable value of $50 billion to $70 billion a year in the United States, from shifting volume to the proper care settings and raising quality of care to reduce emergency room and re-admission rates.

Right value

Open data can help increase value in health-care delivery by controlling costs and improving quality. This involves measures such as providing comparison information to patients so they can select the most cost-effective care, tying provider reimbursement to patient outcomes, and using data to eliminate fraud, waste, or abuse in the system.

Through patient information exchanges that make data available both to providers and patients, patients can become better health-care consumers and help control costs. Through transparency into the cost and quality of different providers and insurance programs, patients can vote with their wallets. There are considerable obstacles to making such information exchanges a reality, including reluctance by providers to share data that could be misinterpreted and the potential difficulty of patients to interpret the data without assistance.

Another source of waste in health care systems is fraud. Fraud in the US Medicare program exceeds $70 billion annually, according to the Government Accountability Office.[64] Regulators can combine data on claims with open data such as demographic information to help identify instances of not providing services that were billed and other fraudulent practices.

Using open and proprietary data to reduce costs, maintain quality, and fight fraud could enable value of $50 billion to $100 billion a year. This would come in the form of reduced costs due by exposing cost and quality data and enabling patients and payers to choose cost-effective treatment plans. Novel payment plans that include performance-based compensation also reduce system costs.

Right innovation

Pharmaceutical companies, device makers, and other medical researchers spend billions of dollars a year developing new therapies, products, and approaches to delivering care. Open data can help make these investments more productive, saving money and bringing benefits to patients sooner.

A key source of value could be in running trials more efficiently by using open and proprietary data to help identify patients to enroll in trials more quickly. This would involve analyzing relevant claims, performing cluster analysis for diseases in particular regions (using public health data, among other data sets), and purchasing social network data. Firms can also use CMS data to find doctors who treat patients with the target disease.

Opening and sharing trial information is not yet the norm, but there are examples; GlaxoSmithKline, for instance, releases patient-level clinical trial data to qualified researchers and providers. This transparency can be helpful for both governments and for companies, which can avoid repeating investments in drugs that have failed. It also can promote higher quality research by making the data available that will allow others to validate results.[65]

---

64 "Medicare and Medicaid fraud, waste, and abuse: Effective implementation of recent laws and agency actions could help reduce improper payments," GAO-11-409T, US Government Accountability Office, March 9, 2011.

65 There has been increasing concern that many published research findings cannot be replicated. See John P. A. Ioannidis, "Why most published research findings are false," *Public Library of Science–Medicine*, volume 2, number 8, August 2005.

The potential value associated with open data in health-care innovation could be $40 billion to $70 billion per year. This would arise from improving the R&D processes of pharmaceutical and medical products companies, including clinical trial operations.

BARRIERS AND RISKS

Health-care systems will have to change substantially for stakeholders to take full advantage of open data. In the United States, traditional levers for controlling costs—such as unit price discounts for procedures based on contracting and negotiating leverage, or elimination of redundant treatments—do not take full advantage of the insights that open data provides. These methods can be augmented or replaced with data- and outcomes-based strategies. Similarly, current approaches to delivering and paying for care, particularly in the United States, will need to be revised. The current system pits payers and providers against each other and focuses decisions on what is and is not covered, rather than on what is and is not most effective. To move beyond the status quo, fee-for-service payment structures would need to be replaced with new systems that base reimbursement on insights on outcomes and costs provided by big data—a move that is already well under way.

One of the largest obstacles will be persuading stakeholders in health care (even patients who stand to gain enormously) to embrace the data-driven approaches that are required to make best use of open data. Both patients and physicians must be willing and able to use insights from the data; this is a personal revolution as much as an analytical one. Doctors, who are trained to make their own inferences, may resist submitting to a regime in which their judgment may be augmented or replaced by judgments based on data from millions of cases. US patients will need to understand and accept that more treatments and tests do not necessarily mean better care.

There also is a risk that open data could be used to perpetuate abuses that exist in current systems. For example, in places where private companies provide diagnostic services, the owners of MRI machines could drive patient volume by using open data to target consumers for treatment based on their characteristics, rather than actual need. Patients, providers, and payers who are interested in "right care" will be wise to watch for such abuses. Open data can also provide information on which to build malpractice suits.

IMPLICATIONS FOR STAKEHOLDERS

Across the health-care ecosystem, open data will bring change and challenges for all stakeholders. Here we examine the implications for all stakeholders and for individual stakeholder groups.

- **All stakeholders.** Today, the words "evidence" and "value" are often defined differently within and across health-care sectors. As a result, payers, providers, and other stakeholders analyze data in different ways. An important foundational step in adopting open data across health care would be to agree on core definitions for evidence and value, as well as developing a consensus about the best analytical protocols.

  All stakeholders will have an interest in and responsibility for dealing with the risks that more liquid data in health care entail. Regulations can be used to clarify how data can be used and how patient privacy is protected and

intellectual property is secured. Given the huge amounts of data involved and the many potential sources of data—patients, providers, payers, government agencies—a high priority may need to be placed on streamlining the process of securing permission to access and use data.[66]

To capture full value from big data, institutions and individuals on the front lines of health care will need to develop capabilities in data analysis, data management, and systems management.

- **Payers.** Payers—insurers and public health systems such as Medicare—have the opportunity to augment their data about patients and providers in their networks with open data. They could use data such as detailed statistics about outcomes and costs both from within and outside their own networks to compare the performance of providers and networks in more detailed ways, identify outliers among providers based on outcomes, and determine the factors that are driving divergent performance. They could also share performance data with clients and members to encourage greater use of the best-performing providers.

  Payers can take several steps to encourage providers to adopt data-driven decision making and effective data capture. They can define value drivers for plan members, as well as the member behaviors and choices that drive value for payers, build clear analytical methods to evaluate expected member value and actual performance, use data to refine communications, and identify resource-intensive workflows and business processes that could be made more efficient through open data. This might include benchmarking processes such as provider authorizations, evaluation of claims accuracy, and auto-adjudication of claims.

  They can also use open and proprietary data to isolate the most important practices that improve the cost of care. For example, they can analyze charges that appear unusual because they deviate from expectations, based on data sets made available by peer organizations. They can track costs of the highest performers, including those related to re-admission, administrative tasks, and laboratory work. And they can compare themselves to metrics that define best-in-class performers, initiating programs to communicate them, and creating incentives to meet these standards.

- **Providers.** Health-care providers have a unique role not only as the primary point of care, but also as a primary source of data origination and capture. There are several ways in which health-care providers can be an effective source for open data that benefits the entire system and can use open and proprietary data to improve their own operations.

  To be an effective source of health-care data, providers can ensure that they capture data consistently and comprehensively and are creating a culture of information sharing to make data available. They can continue to drive adoption and use of EMRs and can reinforce EMR use in patient care. In addition, they can capture data from new sources such as “smart” and

66 Safeguarding patient privacy while simultaneously enabling the value of open data can be challenging. Some experts have argued that the privacy restrictions in the US Health Insurance Portability and Accountability Act (HIPAA) have prevented important types of research. See Association of Academic Health Centers, “HIPAA creating barriers to research and discovery,” June 2008.

embedded medical devices, as well as patient websites, hospital information kiosks, and mobile devices, which can take some physiological readings. Providers can remove technical barriers to sharing information within their organizations and partner with other institutions and third-party providers to share benchmarking data.

Providers can also improve their technology and governance strategies for clinical and operational data. They can establish data ownership and security policies to ensure that organizations have complete access to their own data and understand what the most potent sources of insight are in clinical data. Providers also can create joint clinical and IT teams to define key data needs and cultivate talent with both clinical expertise and advanced data modeling capabilities. In addition, to effectively put open data to use, providers will need to focus on outcomes-based protocols to improve patient care.

- **Pharmaceutical and medical products manufacturers.** Open and proprietary data can help suppliers of medicine and health-care products to meet demands to clearly define the value of their products. This may be a challenging concept, since manufacturers are not typically the source of the "real-world" information after their products enter the market and they have access only to customers who voluntarily self-identify. However, liquid data can be used to establish the total cost of care related to use of their products, as well as the ways in which their products influence patient outcomes. They could develop capabilities that allow them to isolate information related to their products' value and performance, based on open data. They could also build internal governance and investment processes to ensure that the R&D portfolio management process (prioritizing which opportunities to pursue) considers real-world evidence from open data.

  These companies can also use open data to develop a clear view of the efficacy and safety of their products and those of competitors. They could gain access to real-world open data (for example, health-care outcomes after products have been approved and are in wide use) that can give an early indication of possible undetected safety risks. With the right analytical tools and capabilities, companies can use an early warning of trouble to address questions about the efficacy of a product and respond to relevant studies that emerge. Open data can also help companies monitor the safety, efficacy, and value of competing products.

  Finally, these manufacturers could take the lead in making data more liquid. They could start to share clinical trial data (failures and successes) across the R&D community, particularly in high-potential therapeutic areas; this will allow manufacturers to expand their research foundation. To encourage cooperation, the industry could create clear guidelines for intellectual property protection.

- **Government.** Government has a multi-faceted role in the development of open data use in health care. Governments are simultaneously regulators and payers, as well as aggregators of data and users of data. As a provider of data, government in many jurisdictions is committed to making data open.

  Government also has responsibility for making laws and regulations to govern open data and seeing that they are enforced. As a rule maker, government could benefit from a close and ongoing collaboration with other stakeholders to ensure that regulations are strong enough to address the concerns of providers and patients without imposing restrictions that would unduly limit potential benefits. In addition to erecting proper safeguards, government may need to make a substantial financial commitment to enforcement and compliance, making sure that citizens see that violations will be discovered and violators will be punished.

  In its various roles, government has a strong interest in promoting the standards for how open data is collected, aggregated, and formatted to make open data as useful as possible. Government is the natural convener of all stakeholders for such standards setting. Finally, government is the logical stakeholder to take the lead in balancing the evolving concerns of providers, payers, patients, and taxpayers as open data and data-driven approaches to medicine take hold in health care. This may be a role that never ends.

FOR SALE

7. Open data in consumer finance

In this chapter we look at how open data could be applied in three different areas of consumer finance: retail banking, insurance (property and casualty, and life), and residential real estate. Across these industries, open data can be used to improve product offerings, identify new customers, enhance underwriting, and help consumers make better decisions. Given the size and complexity of this domain, we do not attempt to quantify all of the potential value that could be created by using open data levers (unlike in the other domains we analyze in this report), but rather focus on several specific examples that illustrate the impact that open data could have across consumer finance (see Box 11, "Sources of open data in consumer finance").

Box 11. Sources of open data in consumer finance

- **Governments.** Governments provide a wealth of information that is useful in consumer finance, including census and weather data, property records, and geospatial data (detailed maps and GPS data, for example). In addition, governments aggregate consumer complaints about financial products and services as well as data about the performance of investment funds. The US government also provides economic information through the Federal Reserve Economic Data (FRED) service and data about companies filed with the Securities and Exchange Commission through the Edgar system.
- **Financial institutions.** Banks generate a great deal of data based on customer knowledge and transactions, such as consumer preferences and purchasing and savings habits. They also have information about their own products, such as fees, rates, and terms, which can be shared.
- **Insurers.** In the course of underwriting policies and paying claims, insurers collect large amounts of information about consumers. Auto insurers have information about accident rates, types of claims, and driving habits—now, including MyData generated by onboard sensors that drivers agree to use.
- **Credit bureaus.** To calculate credit scores, credit bureaus gather all kinds of financial information, including credit histories, incomes, and assets—for individuals or groups of consumers. These data can be enormously useful for other players in consumer financial services.
- **Third-party developers and data services.** A range of companies aggregate data about financial products and offer consumers comparison data about credit cards, mortgages, student loans, mutual funds, and other products and services.
- **Consumers.** There is a wealth of data about the financial behavior of most consumers. As they shop and pay bills, they generate purchase data, payment histories, and data about use of credit. With sensors and smartphones, consumers are generating information about their habits in new ways.

There is already evidence of the large potential of open data in consumer financial services, as demonstrated by the number of services that aggregate and share data (open and private) about home prices, mortgage rates, credit card offers, individual retirement account (IRA) fees, and other data that enable consumers to make more informed choices. For banks and insurance companies, open data can provide additional information that allows them to assess risk more accurately

and therefore price products more accurately, helping improve margins and also providing a basis for lending to consumers who do not have enough traditional financial data to qualify for a loan.

In this chapter we describe specific levers for value creation in each consumer finance segment.

OPEN DATA IN RETAIL BANKING

The retail banking industry had revenue of $2.3 trillion in 2012.[67] Yet banks are facing margin pressure in the wake of the global financial crisis and consequent reforms aimed at reducing systemic risks that also have depressed returns. This gives the retail banking industry a strong incentive to find more profitable new products and new customers. There is no shortage of potential customers: more than half of the 2.5 billion adults in the world do not use formal financial services to save or borrow. While a majority of these "unbanked" people live in Africa, Asia, and Latin America, 200 million are estimated to live in North America and Europe.[68]

Open data can play a role in helping to serve the unbanked millions. By culling liquid data from sources such as utility bills (if they are shared by utilities), social connections, and census data, banks can gain additional insights into the risk associated with a potential customer—particularly important for assessing consumers who do not have a banking or credit history. Banks can optimize risk management by using open data to screen applications for new services as well as to identify fraud.

Financial institutions are already adopters of big data and are finding ways to apply open data in areas ranging from product design and fraud detection to providing consumers with more information on which to base choices of financial products and providers. We have identified five levers that we believe will be most important in creating value using open data: enhanced product design and pricing, improved consumer marketing, better-informed consumer financial decision making, improved credit-offering decisions, and optimized post-credit decision risk management.

Enhanced product design and pricing

Financial institutions can use open data to improve product design and pricing of traditional products. Examples include using complaint data from regulatory filings to understand what consumers dislike and data collected from social media sites to help identify the most useful credit card features. Emerging peer-to-peer lending platforms are also reliant on open data, often using social media data on an individual's network, along with other public and private information such as education, to pair suitable lenders and borrowers.

Improved consumer marketing

Open data can be used to refine targeting and to customize offerings. Banks can pool internal consumer data and combine it with credit scores from credit bureaus and public data such as information on cost of living and wealth in particular neighborhoods, to more finely segment consumers. This can

67 McKinsey Global Banking Pools.

68 *Half the world is unbanked*, McKinsey & Company and Financial Access Initiative, October 2009.

produce profiles that allow branch employees to suggest additional products for customers, by using data on what consumers in the immediate area purchase. These personalized recommendations, together with an improved sales process, can increase sales conversions by as much as 200 percent.[69]

Better-informed consumer decision making

Open data can help consumers gain a clear view of the costs of increasingly complex financial products. Fee structures and other features of mortgages, credit cards, and other products can be quite opaque, making side-by-side comparisons difficult. Third-party aggregators publish daily updates on rates and fees for mortgages, credit cards, IRAs, and other consumer products, giving consumers up-to-date information on a host of financial products.

This transparency could lead to significant savings, since studies show most consumers do not know the true size of many of these fees: one study found that 93 percent of US consumers underestimated the fees they are paying for the management of their employer-sponsored retirement plans and have no idea how fees are reducing returns.[70] With an estimated $20 trillion of assets under management in pension funds globally, huge amounts of value are at stake.[71] Shifting just 10 percent of the retirement plans that pay the highest fees (about 1.38 percent annually at the 90th percentile) to the average fee (about 0.83 percent) could save $5 billion to $10 billion a year, assuming the funds have the same rate of return.[72] Attaining these savings would depend on employees having access to fee data and employers agreeing to change managers. Third-party services such as BrightScope, an entrepreneurial firm in the United States, help consumers understand how retirement investment plans stack up, both in terms of returns and fees.

Wallaby, a US-based, venture-backed company, is another service that helps customers make better credit decisions by recommending which of their credit cards to use for different types of purchases to maximize rewards. The recommendations are based on open data regarding current offers and promotions for different credit cards.

Another innovative company, BillGuard, aggregates information from all its users on credit card transactions that have been identified as unwanted, erroneous, or even fraudulent. This aggregated information is used to screen all transactions of BillGuard users, thereby highlighting payments that typically are unwanted. The unwanted charges include everything from free trial subscriptions being converted into paid ones without the consumer's noticing, to actual fraud. Investigations show that as many as 0.3 percent of all consumer credit card transactions are erroneous or unwanted and the potential value for consumers globally in identifying and avoiding these payments could be $15 billion to $20 billion a

69 McKinsey analysis.

70 Robert Hiltonsmith, "The retirement savings drain: Hidden and excessive costs of 401(k)s," *Demos*, May 2012. One analysis showed that a couple who invested in a retirement account with a high fee from 1965 to 2005 would have had about $155,000 less at retirement.

71 McKinsey Global Banking Pools.

72 "Inside the structure of defined contribution/401(k) plan fees: A study assessing the mechanics of the 'all-in' fee," Investment Company Institute, 2011. Assuming US employer-sponsored 401(k) plans are representative of global market.

year.[73] This represents the value that would accrue to customers through reduced unwanted charges; banks also could benefit from reduced fraud.

Improved credit-offering decisions

When a bank decides to grant credit, it is putting a price on the risk associated with a particular customer and credit line. To predict the probability of repaying the debt, banks in the United States typically rely on credit histories, which are compiled by credit bureaus that track both positive indicators (staying current on credit card and loan payments) and negative indicators (late payments, defaults). Other economies have similar, if more limited, approaches—tracking only negative indicators or relying on information the bank has aggregated from previous interactions with a particular customer.

The problem with using only credit history is that it leaves a large share of the world's population without access to credit and other financial services. Alternative data sources can provide relevant insights to assess the credit-worthiness of consumers without formal credit histories. Companies such as MicroBilt Corp., which operates a service called Payment Reporting Builds Credit, uses histories of rent, utility, telecom, and other types of bill payments to determine the risk associated with lending to a particular individual. This type of information is beginning to be used by traditional credit bureaus as well; in 2007, Verizon reported the payment history for 20 million of its landline users to TransUnion.[74] In 2010, Experian became the first credit-reporting agency to track tenants' on-time rent payments.

Using open data and other non-traditional data sets to estimate credit-worthiness is a significant opportunity both for banks and consumers. We estimate that the retail banking industry could generate as much as $50 billion to $60 billion per year in additional profits globally by serving about 35 percent of the most profitable of currently unbanked consumers. Using alternate data to qualify customers who currently have access only to expensive credit could unlock an additional potential value of about $40 billion to $50 billion annually in the form of decreased credit costs for the consumers—by enabling a consumer to move to bank credit rather than relying on payday loans, for example.

Optimized post-credit decision risk management

Risk can also be managed after credit is granted, and open data can be used in two ways in this effort. First, banks can detect fraud by screening applications for risk indicators gleaned from a combination of internal data (for example, spending patterns) and external sources of liquid data such as an individual customer's social connections and details of fraud schemes that have been shared by other financial institutions. Banks can also try to minimize the risk of default by creating tailored saving tips based on a consumer's social network or by threatening to publish an individual's financial default to friends and family on social networks, as is currently done by Lenddo, an online community that crowdsources credit scores.[75]

---

73 "The economic impact of grey charges on debit and credit card issuers," Aite Group, July 2013, McKinsey Global Institute analysis. We size only the value for the customer through the reduction in unwanted charges. Banks could benefit as well through fraud reduction.

74 *Extending credit: Helping Americans build solid financial futures*, Brookings Institution, 2007.

75 "Frequently asked questions," Lenddo, www.lenddo.com/pages/faq.

OPEN DATA IN INSURANCE

The life and property and casualty insurance businesses are enormous, but currently slow growing—global revenue in 2012 was about $4 trillion, up only about 1 percent from 2011 in the United States and Europe, reflecting in part the effects of macroeconomic conditions.[76] To generate higher returns, insurers seek to expand the market and lower their cost base, including increasing claims accuracy.

We identify six levers for tapping the value-creating potential of open data in insurance: enhanced product design and pricing, improved consumer marketing, better-informed consumer financial decision making, improved underwriting decisions, improved customer behavior, and enhanced detection of claims abnormalities.

Enhanced product design and pricing

Insurers can refine their product designs (for example, adding new features to life insurance or annuities) by tapping public sources of data, such as consumer complaint data. Innovative use of open data offers the potential to create completely new products such as peer-to-peer insurance products, which rely on a range of data sources such as social media, census data, and geospatial data. One example of this is Friendsurance, a German startup that lets individuals come together and purchase insurance as a group, with members selected based on social network data. These group members pool their premiums to cover the claims costs of the group. Any money left over after all claims have been paid at the end of the year is returned to the group, rather than kept by the insurance company.[77]

Improved consumer marketing

One issue for insurers is the perception that insurance is a commodity product. Insights derived from open data can be used to customize products such as automobile insurance. For example, by using historical weather data, census data, and claims data, a carrier could identify drivers who live in neighborhoods that have experienced damaging hailstorms and develop special offers with hail coverage.

Better-informed consumer financial decision making

Insurance products, particularly life policies and complex annuities, can be very difficult for consumers to understand. There are numerous examples of third-party providers helping consumers find the best insurance for their needs, by collecting open data on prices and terms and then analyzing which insurance policy would have the best rate for that particular consumer. One example of such a service is Ratekick, a venture-backed company in the United States that provides a platform to search insurance rates anonymously, based on insurance company government filings. By anonymously providing seven pieces of information, customers can compare more than 100 insurance providers, based on price and coverage.

76 "Global insurance market report (GIMAR) 2012 edition," International Association of Insurance Supervisors, October 2012; McKinsey Global Insurance Pools.

77 "Friends with benefits," *The Economist*, June 15, 2012.

Improved underwriting decisions

Underwriting—accurately assessing the risk associated with a particular coverage plan for a particular individual and pricing that risk correctly—is a fundamental driver of insurer profitability. Open data are already being used to augment traditional data in this critical function. A number of large US insurance companies have combined internal data sets on customers with open data such as socioeconomic factors, unemployment rates, housing affordability indexes, and average incomes. This information is used to identify consumer behaviors related to increased claims costs. Using predictive analytics to weigh these factors, carriers can price premiums more precisely for a specific customer.

We believe that using open data combined with internal data sets for decision support in underwriting to better select customers and guide them toward certain products has a potential value of $5 billion to $45 billion annually across the globe for the individual property and casualty coverage. This represents a reduction in the combined ratio from 100 percent to 96 percent.[78]

Improved customer behavior

Influencing consumer post-underwriting behavior is a way to lower the risk exposure for insurer and thereby improve claims accuracy. Examples include drivers agreeing to install sensors in their cars to track driving habits, which can qualify them to earn discounts by exhibiting safe driving behavior. This is an example of MyData: Progressive and other insurers that offer the tracking devices provide reports on braking and accelerating habits to participating drivers. Insurance companies use open data, such as weather patterns, climate data, and average distance between home and work, to identify customers whose claims are likely to be escalated (i.e., health and worker compensation payments exceed the initial estimate). Models have shown that drivers commuting long distances in cold and snowy climates are more likely to seek continuation of benefits, possibly because they are concerned about losing control on icy roads when they return to work. Insurance companies use nurses to contact the claimant to ensure the right care is delivered early on so that overall claim accuracy is increased. Some insurers, such as Allianz, are opening their own data and analyses of risk factors (for example, housing types most prone to flooding damage based on geography) and sharing them with customers to let them identify and reduce relevant risk factors.

Enhanced detection of claims abnormalities

An estimated 10 to 15 percent of paid insurance claims are fraudulent, resulting in unnecessary payments that cut into profits and raise overall costs for other customers.[79] Open data can play an important role in screening and flagging suspicious claims. One UK insurer combines internal data on specific claims, risk indicators, and customers with open data, mainly social network data. By uncovering social connections among individuals, the company created a network view of claimants and identified suspicious claims based on connections with known offenders. Social media data have uncovered a number of fraud rings that were composed of individuals making claims against members of the group. Improved fraud detection can reduce claims payouts by as much as 2 to

78 Combined ratio used in the insurance industry to indicate profitability, defined as the sum of claims costs and expenses divided by revenue from premiums.

79 Association of British Insurers.

3 percent, which could mean $15 billion to $30 billion a year across the property and casualty insurance industry catering to individuals.

OPEN DATA IN REAL ESTATE

Open data already plays a big role in some areas of residential real estate—using public records to insure titles, for example—but there are many more areas where open data sources have not yet been applied. We identify five main levers for bringing potential value to developers, home buyers, and lenders: improved location selection for development, improved prioritization of infrastructure investments, optimized seller/leaser and buyer/renter matching, optimized real estate financing, and optimized maintenance planning and execution.

Improved location selection for development

Improving location selection for development can be done using open data sources such as throughput of commuter infrastructure and geospatial data, in tandem with information on sites currently offered for sale. Some developers take this approach to developing both residential and commercial real estate. For example, a developer can use GIS (geographic information system) data along with listings of available properties to identify parcels that meet specific criteria, such as locations within a quarter mile of a light rail station, which might be appealing to a target demographic group in the region.

Improved prioritization of infrastructure investments

Real estate development (and lack of development) has important implications for government. Understanding patterns of real estate development or changes in density is essential for planning infrastructure investments, including roads, water supply lines, schools, sewers, and rail lines. Open data could be used to identify the need for various types of infrastructure investments, such as schools, based on granular demographics information, in combination with economic projections. (Please see the chapter on open data in education for a further discussion of the use of open data in making school location decisions.)

Optimized seller/leaser and buyer/renter matching

Helping potential buyers, renters, leasers, and investors find appropriate properties is critically important to an efficient real estate market. Real estate is usually the largest investment consumers make, raising the stakes for the consumer, who wants to find the right property, at the right price and with the attributes the household needs, such as neighborhood type and proximity to specific amenities. Open data (sales and tax records, mapping data) has already been harnessed by aggregators such as Zillow to show buyers the properties that are for sale, sales histories, and pricing patterns in a given neighborhood, as well as non-financial information such as local infrastructure and location of amenities (the location of the nearest school, for instance). Consumers are then provided a comprehensive overview of property available in different areas and can refine the search with filters for preferences such as walking distance to public transportation. Once a potential property has been identified, the system automatically generates a valuation based on multiple data feeds.

These services provide value to consumers by enabling them to find properties with desired characteristics, a role often played by a broker. It also facilitates transactions, cutting the time properties are on the market and reducing search time for buyers. The value for consumers of these benefits amounts to $20 billion to $30 billion a year in the United States, which represents a 2 to 3 percent reduction in broker fees.

Optimized real estate financing

About $9 trillion in consumer real estate loans are outstanding in the United States.[80] Using open data on income levels, social connections, market price statistics, and purchase history can allow financial institutions to improve their pricing and underwriting of mortgages, enabling them to safely expand their lending. When more information is used to determine risk, consumers with full credit histories may qualify for lower interest rates and consumers with little credit history may gain access the mortgage market. As in the case with retail banking and insurance products, consumers benefit from gaining larger transparency on prices and terms for mortgages with online comparison tools.

We estimate that providing currently underserved consumers access to mortgages through the use of open data sources for credit scoring has a potential value of approximately $10 billion for financial institutions. This increase in profits would be achieved every year by originating loans to 1 percent of households that are under-banked or unbanked. By using alternative data such as a strong record of paying utility bills and meeting other obligations, lenders can extend mortgages to more consumers and save borrowers $25 billion to $30 billion. There is additional value for consumers that currently have a mortgage but can lower their interests cost by using more liquid data to gain a better credit score, but we do not attempt to size this benefit.

Optimized maintenance planning and execution

It is important for all homeowners to maintain the value of their real estate by maintaining it properly. Optimized maintenance planning and execution become more and more relevant with the creation of large distributed real estate operations. Open data on climate and local legislation, and other data sets, could help companies manage maintenance operations more efficiently. For instance, predicting the need for cleanup help before a strong windstorm in a particular area could allow management companies to direct their personnel to the affected regions.

BARRIERS AND ENABLERS

While there are considerable potential benefits for all stakeholders in consumer finance from the use of open data, realizing the value that open data could make possible will require tackling some obstacles.

First, some consumers could resist the use of non-financial data. Poring over phone and utility bills and scanning social media for additional clues about borrower behavior could strike some consumers as an invasion of privacy. They may also have concerns that underwriters could draw negative conclusions about social media activity or associations that are based on incorrect assumptions (e.g., is an applicant for a life insurance policy who spends time on an extreme

80 McKinsey Global Banking Pools.

sports site a participant or simply a fan?). Consumers are already concerned about the difficulty of correcting erroneous information in credit reports from the large credit bureaus that base their scores on actual financial records; correcting non-financial information may seem even more daunting.

While some banks and insurers clearly understand the potential value of tapping open data to enhance underwriting, influence product development, and finely target marketing efforts, these efforts require investments and many institutions do not yet fully appreciate the business case that justifies this cost. Furthermore, the capabilities needed to take advantage of open data across underwriting and marketing functions are not present in many companies. Likewise, easy-to-use tools and reports need to be developed for consumers to capture the value from open data when making financial decisions. Some industries actively try to block the publication of data, since it may reduce their margins. One example is the US real estate business where for many years realtors were hesitant to publish listing data, and still allow only registered realtors to access the Multiple Listing Services.

Furthermore, the legal framework for the use of open data by financial institutions for underwriting and other purposes remains murky. Very little legislation exists on what data can be collected and how it can be used. This is especially challenging for global financial institutions that need to account for many different national policies, and it is seen as the key barrier by many large banks. The lack of clarity has led many institutions to adopt the most conservative policies for all global operations. For example, regulations prohibiting racial and other forms of discrimination in lending have curbed the use of open data by banks because certain types of information may inadvertently correlate with the race, age, and other classes of individuals that are legally protected from discrimination in countries such as the United States.[81] So, while some regional banks in developing economies are using social media as well as other information to create risk models for unbanked consumers, global players do not use such information because it is not explicitly allowed. This lack of clarity may reduce the volume of loans made by global banks.

IMPLICATIONS FOR STAKEHOLDERS

Stakeholders can take specific steps to overcome obstacles to using open data in consumer finance and mitigate the attendant risks.

- **Consumers.** Consumers have a leading role in enabling—or preventing—the use of open data by their banks and insurance companies. They can make efforts to educate themselves and demand clear policies regarding protection of their privacy from financial institutions. In addition, they can work with regulators and elected officials to influence relevant legislation. Consumers can also demand that public sources of relevant financial data be opened up and that easy-to-use applications be available so, for example, they can easily see the volume and severity of consumer complaints about a particular lender.

81 Credit scores have correlations with race, but US law exempts credit scores because they correlate strongly with defaults. It is unclear if regulators would treat additional data the same way they treat credit scores.

- **Financial institutions.** Banks, mortgage companies, and insurers may need to invest in new capabilities to take full advantage of the open data opportunity. Also, it will be largely up to these institutions to overcome potential consumer resistance by educating the public about the benefits of open data and sharing their methodologies and policies. It may also be incumbent upon financial institutions to lay out standards for data use in each region and engage legislators in formalizing policies in each country. Companies can share the open data they collect on consumers as MyData and make it easy for consumers to correct erroneous data.

- **Third-party vendors.** Third-party data brokers and developers can play an important role in making open financial data easy to access and understand, helping to spur use (for example, for comparison shopping) and to prove the value of open data to consumers. A great deal of work is yet to be done in perfecting comparison engines and creating useful visualizations. For financial institutions, third-party data sources and tools will help accelerate use. In addition, third-party vendors can help create standards that will work for multiple stakeholders.

- **Governments.** Government acts as a source of open data and will likely be the rule makers for many uses of open data in consumer finance, too. The highest priority will be to establish the rules that are needed both to protect the rights of individuals and to provide the clear regulatory foundation financial institutions need before they can fully capture the potential of open data. Legislators can facilitate the public conversation about trade-offs between privacy and the benefits of open data. Finally, government can provide the mechanisms for consumers and businesses to seek redress and compensation for misused data and other abuses.

Bibliography

Aite Group, "The economic impact of grey charges on debit and credit card issuers," 2013.

American Public Transportation Association, "Rising gas prices equal big savings for public transportation riders," press release, July 19, 2013.

Arizona State University, *Annual report*, www.annualreport.asu.edu.

Association of Academic Health Centers, "HIPAA creating barriers to research and discovery," June 2008.

BBC World News, "Tim Berners-Lee unveils government data project," January 21, 2010.

Beckman, Jeremy, "Statoil aims to squeeze more out of Oseberg," *Offshore*, October 1, 2012.

Berkowitz, Elana, and Renée Paradise, "Innovation in government: Kenya and Georgia," *McKinsey Quarterly*, September 2011.

Biswas, Shreya, "Attrition in India to top world charts in 2013; one in four employees to change jobs," *The Economic Times*, June 7, 2013.

Boston School Choice, http://bostonschoolchoice.org.

Brookings Institution, *Extending credit: Helping Americans build solid financial futures*, January 7, 2007.

Center for American Progress, *Size matters: A look at school-district consolidation*, August 2013.

College Abacus, www.collegeabacus.com.

Consumer Product Safety Commission, www.cpsc.gov.

Cruz, Marcos, and Alexandre Lazarow, "Innovation in government: Brazil," *McKinsey Quarterly*, September 2012.

The Economist, "Friends with benefits," June 15, 2012.

Education Growth Advisors, *Learning to adapt: Understanding the adaptive learning supplier landscape*, Bill and Melinda Gates Foundation, 2013.

Education Superhighway, "Procurement gap," www.educationsuperhighway.org.

Energy Information Agency, *Annual energy outlook 2013*, May 2013.

Energy Star, *Energy Star products: 20 years of helping America save energy, save money, and protect the environment*, 2012.

Federal Highway Administration, *Good practice in value pricing*, May 22, 2012.

Friedman Foundation for Educational Choice, "How does school choice work in other countries?" August 2013.

Gigler, S., S. Custer, and H. Rahemtulla, *Realizing the vision of open government* data*: Opportunities, challenges and pitfalls*, World Bank Open Development Technology Alliance working paper series number 3, 2011.

Gomez, Miguel, "Coffee market overview: Trends and consumer preferences," presented at Let's Talk Coffee conference in Antioquia, Colombia, October 4, 2012.

(US) Government Accountability Office. "Medicare and Medicaid fraud, waste, and abuse: Effective implementation of recent laws and agency actions could help reduce improper payments," *GAO-11-409T*, March 9, 2011.

Hanushek, Eric, "Valuing teachers: How much is a good teacher worth?" *Education Next*, volume 11, number 3, Summer 2011.

Headlee, Jim, "Reverse auctions revolutionize purchasing," *Illinois School Board Journal*, May/June 2010.

Hiltonsmith, Robert, "The retirement savings drain: Hidden and excess costs of 401(k)s," *Demos*, May 2012.

International Association of Insurance Supervisors, "Global insurance market report (GIMAR) 2012 edition," October 2012.

Investment Company Institute, *Inside the structure of defined contribution/401(k) plan fees: A study assessing the mechanics of the 'all-in' fee*, November 2011.

Ioannidis, John P. A., "Why most published research findings are false," *Public Library of Science–Medicine*, volume 2, number 8, August 2005.

Kaplan, Siena, Pedro Morillas, Emily Rusch, and Phineas Baxandall, "California budget transparency 2.0: Online tools for better government," *CALPIRG Education Fund*, January 2009.

Kundra, Vivek, Data.*gov: Pretty advanced for a one-year-old*, whitehouse.gov, May 21, 2010.

McKinsey & Company, *The "big data" revolution in health care: Accelerating value and innovation*, January 2013.

McKinsey & Company, *Impact of the financial crisis on carbon economics: Version 2.1 of the global greenhouse gas abatement cost curve*, 2010.

McKinsey & Company Electric Power and Natural Gas Practice and Business Technology Office, *US smart grid value at stake: The $130 billion question*, 2009.

McKinsey & Company and Financial Access Initiative, *Half the world is unbanked*, March 2009.

McKinsey Center for Government, *Education to employment: Designing a system that works*, December 2012.

McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity*, May 2011.

McKinsey Global Institute, *Disruptive technologies: Advances that will transform life, business, and the global economy*, May 2013.

OECD, *Education at a glance*, 2010, 2011.

Open Government Working Group, "Eight principles of open government data," opengovdata.org.

Park, Kristen, and Miguel I. Gomez, "Do price premiums exist for local products?" *Journal of Food Distribution Research*, volume 43, number 1, March 2012.

Park, Sophie, "We need entrepreneurs to be given the opportunity to promote their tools and services to anyone who is interested," interview with Tim Kelsey, NHS National Director for Patients and Information, Health 2.0 News, www.health2con.com, August 8, 2013.

Phillips, Matthew, and Asjylyn Loder, "Amid US oil boom, railroads are beating pipelines in crude transport," *BloombergBusinessweek*, June 13, 2013.

Rahemtulla, H., S. Custer, I. Risacova, K. Jhalla, S. Gigler, and C. Brigham, *The journey of open government and open* data *Moldova*, World Bank Publications, July 9, 2012.

Save energy at home 2012, www.energystar.gov.

Seattle Department of Transportation, *Seattle variable tolling study*, May 2009.

Snyder, Thomas, *120 years of American education: A statistical portrait*, National Center for Education Statistics, US Department of Education, 1993.

Stone, Brad, "Why Redfin, Zillow, and Trulia haven't killed off real estate brokers," *BloombergBusinessweek*, March 7, 2013.

"Submarine piloting and navigation training," Office of Naval Research, www.onr.navy.mil/en.aspx.

Texas Transportation Institute, *Selecting travel reliability measures*, 2003.

Tischler, Linda, "He struck gold on the Net (really)," *Fast Company*, June 2002.

United Kingdom Cabinet Office, *Open* data *white paper: Unleashing the potential*, June 28, 2012.

United Kingdom Cabinet Office, *Shakespeare review: An independent review of public sector information*, May 2013.

Vaznis, James, "MIT has plan for Boston school assignments," *The Boston Globe*, October 28, 2012.

Viswanathan, Meera, Carol F. Golin, Christine E. Jones, Mahima Ashok, Susan J. Blalock, Roberta C. M. Wines, Emmanuel J. L. Coker-Schwimmer, David L. Rosen, Priyanka Sista, and Kathleen N. Lohr, "Interventions to improve adherence to self-administered medications for chronic diseases in the United States: A systematic review," *Annals of Internal Medicine*, volume 157, number 11, December 4, 2012.

Wendy Schmidt Oil Cleanup X Challenge, www.iprizecleanoceans.org.

Werdegier, Julie, "Tesco, British grocer, uses weather data to predict sales," *New York Times*, September 1, 2009.

Where the money goes: Window on state government, www.texastransparency.org.

Yam, Phillip, "How to kick-start innovation with free data," *Scientific American*, March 23, 2013.

Related McKinsey Global Institute research

McKinsey Insights iPad app (September 2013)

This new app provides mobile access to the latest perspectives from McKinsey, MGI, and the *McKinsey Quarterly*. Our articles and reports address the most challenging issues facing senior leaders around the world, spanning countries, industries, and all business functions to examine leadership and corporate strategy to organization, technology, marketing, and operations. The app is available from the Apple App Store.

Disruptive technologies: Advances that will transform life, business, and the global economy (May 2013)

It is critical that business and policy leaders understand which technologies will matter to them and prepare accordingly. This reports cuts through the noise and identifies 12 technologies that could drive truly massive economic transformations and disruptions in the coming years. It also looks at exactly how these technologies could change our world, as well as their benefits and challenges, and offers guidelines to help leaders from businesses and other institutions respond.

The social economy: Unlocking value and productivity through social technologies (July 2012)

MGI analyzes the potential impact of social technologies. Based on in-depth analyses of how social technology can be used in five economic sectors (four commercial sectors, plus the social sector), the report identifies a series of value levers that can be applied across the value chain and within and across enterprises.

Online and upcoming: The Internet's impact on aspiring countries (January 2012)

This report explains how the Internet today connects about two billion people worldwide. Half of these are in the "aspiring" world—countries as varied as Algeria, South Africa, China, Iran, and Mexico—that are climbing the developmental ladder quickly, with diverse populations and inarguable economic potentialities. It examines the impact of the Internet in populous and fast-growing aspiring countries, where it offers even greater potential than in the developed world.

Big data: The next frontier for innovation, competition, and productivity (May 2011)

Big data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus—as long as the right policies and enablers are in place.

www.mckinsey.com/mgi

E-book versions of selected MGI reports are available at MGI's website, Amazon's Kindle bookstore, and Apple's iBookstore.

Download and listen to MGI podcasts on iTunes or at www.mckinsey.com/mgi/publications/multimedia/

McKinsey Global Institute
October 2013
Copyright © McKinsey & Company
www.mckinsey.com/mgi
@McKinsey_MGI
McKinseyGlobalInstitute